(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 January 2003 (03.01.2003)

PCT

(10) International Publication Number
WO 03/000715 A1

(51) International Patent Classification[7]: C07K 1/00, 14/00, 17/00, A01K 67/00, 67/033, G01N 33/00, 33/53, 33/567, C12Q 1/00, C12N 15/63, 15/85, 15/87

(21) International Application Number: PCT/US02/19750

(22) International Filing Date: 21 June 2002 (21.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/300,135 22 June 2001 (22.06.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/300,135 (CIP)
Filed on 22 June 2001 (22.06.2001)

(71) Applicant (*for all designated States except US*): **CERES, INC.** [US/US]; 3007 Malibu Canyon Road, Malibu, CA 90265 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **DANG, Van-Dinh** [VN/US]; 274 Saint Criox Court, Oak Park, CA 91377 (US). **OKAMURO, Jack** [US/US]; 307 Ocho Rios Way, Oak Park, CA 91377 (US).

(74) Agent: **LUNDQUIST, Ronald, C.**; Fish & Richardson P.C., P.A., Suite 3300, 60 South Sixth Street, Minneapolis, MN 55402 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: CHIMERIC HISTONE ACETYLTRANSFERASE POLYPEPTIDES

(57) Abstract: Chimeric polypeptides are disclosed that comprise a first polypeptide segment having histone acetyltransferase enzymatic activity and a second polypeptide segment that is similar to a subunit of a chromatin-associated histone deacetyltransferase protein complex. Also disclosed are nucleic acids encoding such chimeric polypeptides and eukaryotic organisms expressing such chimeric polypeptides.

WO 03/000715 A1

Chimeric Histone Acetyltransferase Polypeptides

TECHNICAL FIELD

This invention relates to methods and materials for analyzing and modulating gene expression. In particular, the invention features chimeric histone acetyltransferase polypeptides that can be used to determine gene expression profiles in specific cells, and to modulate gene expression in specific cells.

BACKGROUND

Genes often are differentially expressed during the development of an organism, and in particular cells in an organism. Understanding and manipulating an organism's temporal and spatial gene expression profile can be useful for developing new and improved biological products and therapies. Among the array of regulatory mechanisms that affect the gene expression profile of an organism, chromatin remodeling has an important role.

Eukaryotic DNA is tightly packaged into chromatin. The most basic element of DNA packaging is the nucleosome, which consists of an octamer of histone proteins wrapped by about 146 nucleotide base pairs. The compaction of eukaryotic DNA into nucleosomes and the formation of nucleosome arrays present natural barriers to genetic regulatory proteins, and to enzymes that interact with DNA. Chromatin-associated protein complexes reportedly can, among other things, stabilize and destabilize nucleosomal DNA and thereby affect nuclear processes that use DNA as a substrate (e.g., transcription, replication, DNA repair, and DNA organization) as well as regulators of these processes.

Some chromatin-associated protein complexes are reported to use the energy of ATP hydrolysis to increase histone mobility, and to thereby change the accessibility of certain nucleosomal DNA to enzymes that process genetic information and to genetic regulatory proteins. It is thought that ATP-dependent chromatin-remodeling protein complexes can have a role in both gene activation and repression. Researchers have reported the existence of ATP-dependent chromatin-remodeling protein complexes in organisms including yeast (e.g., SWI/SNF, RSC, ISW1, ISW2, and Ino 80), Drosophila

(e.g., dSWI/SNF, ACF, CHRAC, and NURF), and human (e.g., hSWI/SNF, NuRD, RSF, and ACF).

Other chromatin-associated protein complexes are reported to change chromatin structure by covalently modifying histones (e.g., by adding or removing acetyl, methyl, phosphate or ubiquitin). It is thought that by covalently modifying histones, these protein complexes can affect chromatin structure and thereby change the accessibility of nucleosomal DNA to enzymes that process genetic information and to genetic regulatory proteins. Some of these histone-modifying protein complexes also are thought to affect the activity of ATP-dependent chromatin-remodeling complexes.

For example, some histone-modifying chromatin-associated protein complexes reportedly contain a polypeptide subunit having histone acetyltransferase ("HAT") enzymatic activity. Such protein complexes are, in general, thought to have a role in activating transcription. Researchers have reported the existence of polypeptides having HAT enzymatic activity in organisms including yeast, *Tetrahymena*, and humans.

As another example, some histone-modifying chromatin-associated protein complexes reportedly contain a polypeptide subunit having histone deacetylase ("HDAC") enzymatic activity. Such protein complexes are, in general, thought to have a role in repressing transcription. Researchers have reported the existence of polypeptides having HDAC enzymatic activity in organisms including yeast, *C. elegans*, *Drosophila*, *Xenopus*, chicken, mouse, human and maize.

SUMMARY

The present invention relates to chimeric histone acetyltransferase ("HAT") polypeptides useful for determining gene expression profiles in specific cell types, or for modulating gene expression in specific cell types. For example, chimeric HAT polypeptides can be used to affect gene expression to achieve desirable results, such as enhancing expression of specific genes in a eukaryotic organism. Chimeric HAT polypeptides contain a polypeptide segment that has HAT enzymatic activity and a polypeptide segment that is similar or identical to a subunit a of chromatin-associated protein complex having histone deacetylase ("HDAC") enzymatic activity.

Thus, the invention features chimeric polypeptides that contain: 1) a first polypeptide segment that exhibits histone acetyltransferase activity, and 2) a second polypeptide segment having 40% or greater (e.g., at least 40%, at least 60%, at least 80% and at least 90%) sequence identity to a subunit of a histone deacetylase chromatin-associated protein complex (e.g., a subunit that exhibits scaffold activity, a subunit that exhibits DNA binding activity, a subunit that exhibits ATPase-dependent helicase activity, and a subunit that exhibits histone deacetylase activity). The first and second polypeptide segments are arranged such that a terminus of the second polypeptide segment is linked to a terminus of the first polypeptide segment via at least one covalent bond.

In some embodiments, the first and second polypeptide segments can be directly linked via a peptide bond. In such embodiments the C-terminal amino acid of the first polypeptide segment can be linked to the N-terminal amino acid of the second polypeptide segment. Alternatively, the N-terminal amino acid of the first polypeptide segment can be linked to the C-terminal amino acid of the second polypeptide segment. In some embodiments, the first and second polypeptide segments can be indirectly linked via one or more (e.g., 1 to 50, and 10 to 50) intervening amino acids that are situated between the first and second polypeptides. In such embodiments, the C-terminal amino acid of the first polypeptide segment can be linked to an intervening amino acid, and the N-terminal amino acid of the second polypeptide segment can be linked to an intervening amino acid. Alternatively, the N-terminal amino acid of the first polypeptide segment can be linked to an intervening amino acid, and the C-terminal amino acid of the second polypeptide segment can be linked to an intervening amino acid. In some embodiments, the intervening amino acids include at least one alanine residue and / or at least one glycine residue.

The invention also features nucleic acid constructs that encode such chimeric polypeptides, and eukaryotic organisms that include such chimeric polypeptides.

The invention also features eukaryotic organisms that contain a nucleic acid that encodes a chimeric polypeptide having: 1) a first polypeptide segment that exhibits histone acetyltransferase activity; and 2) a second polypeptide segment that has 40% or greater sequence identity to a subunit of a histone deacetylase chromatin-associated

protein complex. The first and second polypeptide segments of the encoded chimeric polypeptide are arranged such that a terminus of the second polypeptide segment is covalently linked to a terminus of the first polypeptide segment. The nucleic acid can be operably linked to a promoter.

The invention also features eukaryotic organisms that contain: 1) a first nucleic acid construct having a first promoter and a transcription activator element operably linked to a coding sequence that encodes a chimeric polypeptide, and 2) a second nucleic acid construct having a second promoter conferring cell type-specific transcription operably linked to a coding sequence for a polypeptide that binds the transcription activator element. The encoded chimeric polypeptide has: 1) a first polypeptide segment that exhibits histone acetyltransferase activity, and 2) a second polypeptide segment that has 40% or greater sequence identity to a subunit of a histone deacetylase chromatin-associated protein complex. The first and second polypeptide segments of an encoded chimeric polypeptide are arranged such that a terminus of the second polypeptide segment is covalently linked to a terminus of the first polypeptide segment. In some embodiments, the organism is an animal. In other embodiments the organism is a plant (e.g., a monocot such as corn and rice, or a dicot such as soybean and rape). In some embodiments, the plant contains a mutation or agent that alters (i.e., increases or decreases) the DNA methylation state in the plant relative to a corresponding plant that lacks said agent or mutation. In some embodiments, the mutation is in a C5 DNA methyltransferase (a.k.a. cytosine C5 DNA methyltransferase) gene. In some embodiments, the agent is an antisense nucleic acid. In some embodiments, the agent affects expression of a C5 DNA methyltransferase gene.

The invention also features methods for detecting the expression of one or more genes in a eukaryote. The methods involve isolating macromolecules from one or more specific cells in a eukaryote (e.g., a plant or an animal) that contains a nucleic acid construct in which a promoter is operably linked to a coding sequence that encodes a chimeric polypeptide, and then determining the presence or amount of at least one of the macromolecules in at least one of the specific cells. The encoded chimeric polypeptide has: 1) a first polypeptide segment that exhibits histone acetyltransferase activity, and 2) a second polypeptide segment that has 40% or greater sequence identity to a subunit of a

histone deacetylase chromatin-associated protein complex. The first and second polypeptide segments of the encoded chimeric polypeptide are arranged such that a terminus of the second polypeptide segment is covalently linked to a terminus of the first polypeptide segment. In some embodiments, the macromolecules are polypeptides. In some embodiments, the macromolecules are nucleic acids. In some embodiments, the promoter confers cell-type specific transcription in a plant reproductive tissue (e.g., ovule, central cell, endosperm, embryo, and zygote). In some embodiments, the promoter confers cell-type specific transcription in a plant vegetative tissue.

In some embodiments, the eukaryote also contains a second nucleic acid construct. In such embodiments, the first nucleic acid construct has a recognition site for a transcriptional activator operably linked to the promoter and the coding sequence. The second nucleic acid construct has a second promoter conferring cell-type specific transcription that is operably linked to a coding sequence for a polypeptide that binds the recognition site for the transcriptional activator.

The invention also features methods for modulating gene expression in a eukaryote. The methods involve making a eukaryote (e.g., a plant or an animal) having a nucleic acid construct in which a cell-type specific promoter is operably linked to a coding sequence that encodes a chimeric polypeptide. The encoded chimeric polypeptide has: 1) a first polypeptide segment that exhibits histone acetyltransferase activity, and 2) a second polypeptide segment that has 40% or greater sequence identity to a subunit of a histone deacetylase chromatin-associated protein complex. The first and second polypeptide segments of the encoded chimeric polypeptide are arranged such that a terminus of the second polypeptide segment is covalently linked to a terminus of the first polypeptide segment. The eukaryote exhibits modulated gene expression in cells in which the promoter confers cell-type specific transcription. In some embodiments, the eukaryote has compositional alterations relative to a corresponding organism that lacks said nucleic acid construct. In some embodiments, the eukaryote has developmental alterations relative to a corresponding organism that lacks said nucleic acid construct. In some embodiments, the eukaryote has phenotypic alterations relative to a corresponding organism that lacks said nucleic acid construct.

In some embodiments, the organism is a plant. In some embodiments, the promoter confers cell-type specific transcription in a plant reproductive tissue (e.g., ovule, central cell, endosperm, embryo, and zygote). In some embodiments, the promoter confers cell-type specific transcription in a plant vegetative tissue. In some embodiments, the plant contains a mutation or agent that alters (e.g., increases or decreases) the DNA methylation state in the plant relative to a corresponding plant that lacks said agent or mutation. In some embodiments, the mutation is in a C5 DNA methyltransferase gene. In some embodiments, the agent is an antisense nucleic acid. In some embodiments, the agent affects expression of a C5 DNA methyltransferase gene. In some embodiments, modulated gene expression alters seed development. In some embodiments modulated gene expression alters embryo development. In some embodiments, modulated gene expression alters endosperm development. In some embodiments, modulated gene expression alters seed yield by mass.

The invention also features methods for modulating gene expression in a eukaryote that involve making a eukaryote (e.g., a plant or an animal) that has 1) a first nucleic acid construct having a first promoter and a transcription activator element operably linked to a coding sequence that encodes a chimeric polypeptide, and 2) a second nucleic acid construct having a second promoter conferring cell type-specific transcription operably linked to a coding sequence for a polypeptide that binds the transcription activator element. The encoded chimeric polypeptide has: 1) a first polypeptide segment that exhibits histone acetyltransferase activity, and 2) a second polypeptide segment that has 40% or greater sequence identity to a subunit of a histone deacetylase chromatin-associated protein complex. The first and second polypeptide segments of the encoded chimeric polypeptide are arranged such that a terminus of the second polypeptide segment is covalently linked to a terminus of the first polypeptide segment. The eukaryote exhibits modulated gene expression in cells in which the second promoter confers cell-type specific transcription. In some embodiments, the eukaryote has compositional alterations relative to a corresponding organism that lacks said nucleic acid construct. In some embodiments, the eukaryote has developmental alterations relative to a corresponding organism that lacks said nucleic acid construct. In some

embodiments, the eukaryote has phenotypic alterations relative to a corresponding organism that lacks said nucleic acid construct.

In some embodiments, the organism is a plant. In some embodiments, the second promoter confers cell-type specific transcription in a plant reproductive tissue (e.g., ovule, central cell, endosperm, embryo, and zygote). In some embodiments, the second promoter confers cell-type specific transcription in a plant vegetative tissue. In some embodiments, the plant contains a mutation or agent that alters (e.g., increases or decreases) the DNA methylation state in the plant relative to a corresponding plant that lacks said agent or mutation. In some embodiments, the mutation is in a C5 DNA methyltransferase gene. In some embodiments, the agent is an antisense nucleic acid. In some embodiments, the agent affects expression of a C5 DNA methyltransferase gene. In some embodiments, modulated gene expression alters seed development. In some embodiments modulated gene expression alters embryo development. In some embodiments, modulated gene expression alters endosperm development. In some embodiments, modulated gene expression alters seed yield by mass.

The invention also features methods for making a genetically modified eukaryote. The methods involve making a first eukaryote (e.g., a plant or an animal) that has a first nucleic acid construct having a first promoter and a transcription activator element operably linked to a coding sequence. The coding sequence encodes a first polypeptide segment and a second polypeptide segment. The first polypeptide segment exhibits histone acetyltransferase activity, and the second polypeptide segment has 40% or greater sequence substantially identical to a subunit of a histone deacetylase chromatin-associated protein complex. The first and second polypeptide segments of the encoded chimeric polypeptide are arranged such that a terminus of the second polypeptide segment is covalently linked to a terminus of the first polypeptide segment. The methods also involve making a second eukaryote that has a second nucleic acid construct having a promoter that confers embryo-specific transcription operably linked to a coding sequence encoding a polypeptide that binds the transcription activator element of the first nucleic acid construct. The method also involves crossing the first and second eukaryotes to form genetically modified progeny that are sterile.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used to practice the invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description.

DETAILED DESCRIPTION

This invention features chimeric histone acetyltransferase ("HAT") polypeptides. Chimeric HAT polypeptides can be used to determine and modulate gene expression profiles in eukaryotic organisms.

Chimeric polypeptides

A chimeric HAT polypeptide contains at least two polypeptide segments: a first polypeptide segment that exhibits HAT enzymatic activity, and a second polypeptide segment that is substantially identical to a subunit of those chromatin-associated protein complexes having histone deacetyltransferase ("HDAC") activity. A chimeric HAT polypeptide typically is not found in nature.

First polypeptide segment

A polypeptide segment that exhibits HAT enzymatic activity is a suitable first polypeptide segment of a chimeric HAT polypeptide. Whether a first polypeptide segment exhibits HAT enzymatic activity can be determined by testing either the polypeptide segment or the chimeric HAT polypeptide using an assay that measures the transfer of an acetyl functional group from an acetyl donor such as acetyl CoA to a histone polypeptide or polypeptide segment. See e.g., Brownell, J. and Allis, C.D. (1995) *Proc. Natl. Acad. Sci.* 92, 6364-6368. This assay can be used to screen candidate

polypeptide segments for HAT enzymatic activity, and to test chimeric polypeptides for HAT enzymatic activity.

In some embodiments, a first polypeptide segment has an amino acid sequence that corresponds to the amino acid sequence of one of the following polypeptides: yeast Esa1, Gcn5, Sas3, yTAFIII130, ELP3, HAT1 or Hpa2; *Drosophila* dGcn5, dTAFII230 or MOF; *Tetrahymena* p55; or human hGcn5, p300/CPB, PCAF, Tip60, hTAFII250, TFIII90/110/220, SRC-1 or ACTR. In other embodiments, a first polypeptide segment can have an amino acid sequence with substitutions, insertions or deletions relative to one of the above-mentioned polypeptides. Any polypeptide segment having HAT enzymatic activity is suitable as a first polypeptide segment, irrespective of the number or character of amino acid insertions, deletions, or substitutions. Thus, in some embodiments, the amino acid sequence of a first polypeptide segment corresponds to less than the full-length sequence (e.g., a HAT functional domain) of one of the above-mentioned polypeptides.

One of skill will recognize that individual substitutions, deletions or additions to a polypeptide that alter, add or delete a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. The following six groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Serine (S), Threonine (T);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

(see e.g., Creighton, *Proteins* (1984)).

Other suitable candidates for first polypeptide segments can be identified by homologous polypeptide sequence analysis. A similar analysis can be applied to identify suitable candidates for second polypeptide segments. HAT amino acid sequence families

are known to be conserved. For example, plant histone acetyltransferase genes can be identified by BLAST or PSI-BLAST analysis of nonredundant protein databases using known plant, yeast and/or animal histone acetyltransferase amino acid sequences. Homologous polypeptide sequence analysis involves the identification of conserved regions in a template polypeptide, also referred to herein as a subject polypeptide. Conserved regions can be identified by locating a region within the primary amino acid sequence of a template polypeptide that is a repeated sequence, forms some secondary structure such as helices, beta sheets, etc., establishes positively or negatively charged domains and represents a protein motif or domain. See e.g., Bouckaert et al., U.S. Ser. No. 60/121,700, filed February 25, 1999, and the Pfam web site describing consensus sequences for a variety of protein motifs and domains at http://www.sanger.ac.uk/Pfam/ and http://genome.wustl.edu/Pfam/. The information included in the Pfam database is described in Sonnhammer et al., *Nucl Acids Res* 26:320-322 (1998), and in Sonnhammer et al., *Proteins* 28:405-420 (1997); Bateman et al., *Nucl Acids Res* 27:260-262 (1999), and Sonnhammer et al., *Proteins* 28:405-20 (1997). From the Pfam database, consensus sequences of protein motifs and domains can be aligned with the template polypeptide sequence to determine conserved region(s).

Conserved regions also can be determined by aligning sequences of the same or related polypeptides from closely related plant species. Closely related plant species preferably are from the same family. Alternatively, alignment are performed using sequences from plant species that are all monocots or are all dicots. In some embodiments, alignment of sequences from two different plant species is adequate. For example, sequences from canola and *Arabidopsis* can be used to identify one or more conserved regions. Such related polypeptides from different plant species need not exhibit an extremely high sequence identity to aid in determining conserved regions. For example, polypeptides that exhibit about 35% sequence identity can be useful to identify a conserved region. Typically, conserved regions of related proteins exhibit at least 40% sequence identity; or at least about 50%; or at least 60%, or at least 70%, at least 80%, or at least 90% sequence identity. In some embodiments, a conserved region of target and template polypeptides exhibit at least 92, 94, 96, 98, or 99% sequence identity. Sequence identity can be either at the amino acid or nucleotide level.

In some embodiments, a first polypeptide segment is the polypeptide encoded by the maize HAC101, HAC104, HAC105, HAC107 or HAC109 gene or a homolog thereof. The maize HAC101 gene belongs to the CREB-Binding Protein family of transcriptional co-activators with histone acetyltransferase activity. Maize HAC104 is most homologous to the GCN5 family of HATs in yeast and animals. Maize HAC105 is most homologous to the ESA1 related family of HATs in yeast and animals. Maize HAC107 is most homologous to the ELP3 related family of HATs in yeast and animals. Maize HAC109 is most homologous to the HAT1 related family of HATs in yeast and animals. In other embodiments, polypeptides having modifications relative to the above polypeptides are suitable first polypeptide segments.

In some embodiments, a first polypeptide segment is the polypeptide encoded by the *Arabidopsis* HAC1, HAC2, HAC3, HAC4, HAC7 or HAC8 gene or a homolog thereof. *Arabidopsis* HAC2 and HAC4 genes encode HATs that are homologous to human CREB-binding proteins. *Arabidopsis* HAC3 is homologous to yeast Gcn5. *Arabidopsis* HAC1 is homologous to yeast HAT1. In other embodiments, polypeptides having modifications relative to the above polypeptides are suitable first polypeptide segments.

Exemplary amino acid sequences of HAT polypeptides are shown in Table 6.

Yet other first polypeptide segments can be synthesized on the basis of consensus HAT functional domains. See e.g., Table 13.

Second polypeptide segment

Chimeric polypeptides of the invention have a second polypeptide segment that is covalently linked to the first polypeptide segment. A second polypeptide segment can have substantial identity, or can be identical, to a subunit of certain chromatin-associated protein ("CAP") complexes, i.e., those CAP complexes having a subunit that exhibits histone deacetylase activity ("CAP/HDAC complexes"). CAP/HDAC complexes include, for example, polycomb group (PcG) complexes, SIN3/HDAC-containing complexes, Mad-Max complexes, Tup1-Ssn6 complexes, DNMT1 complexes, MeCP1 and MeCP2 complexes, MBD complexes, and Ikaros-Aiolos-containing complexes.

Amino acid sequences of subunits of CAP/HDAC complexes generally are conserved among different species.

CAP/HDAC complexes can be distinguished from other chromatin-associated protein complexes by the presence of a subunit that exhibits histone deacetylase activity. Alternatively, CAP/HDAC complexes can be distinguished from other chromatin-associated protein complexes by the presence of a subunit that exhibits sequence homology to known histone deacetylase proteins. In contrast, other chromatin-associated protein complexes either have histone acetyltransferase activity or have neither HAT nor HDAC activity. CAP/HDAC complexes also can be distinguished from other chromatin-associated protein complexes by their effect, *in vitro* or *in vivo*, on gene expression. Transcription from genes in nucleosomes to which CAP/HDAC complexes are bound typically is reduced or even eliminated. In contrast, chromatin-associated protein complexes having a HAT subunit typically facilitate increased transcription from genes in nucleosomes to which such complexes are bound. CAP/HDAC complexes can be distinguished from transcription complexes by the lack of any subunit that interacts directly with RNA polymerase II. CAP/HDAC complexes can be readily distinguished from nucleosomes because CAP/HDAC complexes do not have histones as subunits of the complex.

Whether a particular complex possesses a subunit that exhibits HDAC activity can be determined by testing a putative CAP/HDAC complex or its subunits, for HDAC activity. HDAC activity can be determined by an assay that measures the removal of an acetyl group from a histone polypeptide or histone polypeptide segment. See e.g., van der Vlag, J. and Otte, A.P. *Nature Genetics* 25, 474-478 (1999). This assay can be used to screen subunits of candidate CAP complexes for HDAC activity. Alternatively, a CAP complex can be shown to possess a subunit having HDAC activity by sequence identity to a subunit of a known CAP/HDAC complex, as described herein.

Once a CAP complex has been determined to possess a histone deacetylase as one subunit of the complex, then all subunits of that particular CAP/HDAC complex can be tested for their suitability as a second polypeptide segment. Polypeptides can be identified as subunits of a CAP/HDAC complex by their co-purification with the complex.

In some embodiments, the second polypeptide segment is the subunit that is HDAC itself. Such subunits can be identified using the above-described assay for HDAC enzymatic activity. The following polypeptides having HDAC enzymatic activity have been identified: yeast RPD3, HDA1, HOS1, HOS2, and HOS3; *C. elegans* HDA1, HDA2, HDA3; *Drosophila* dHDAC1, dHDAC2, dHDAC3, and dHDA2; *Xenopus* HDm; chicken HDAC1, HDAC2, and HDAC3; mouse HDAC1, HDAC2, HDAC3, mHDA1, and mHDA2; human HDAC1, HDAC2, HDAC3, HDAC4, HDAC5, HDAC6, HDAC7, and HDAC8; and maize RPD3 and HD2. See e.g., Cress W.D. and Seto, E. (2000) *J. Cell. Physiol.* 184, 1-16. All of the above HDAC polypeptides are suitable for use as the second polypeptide segment, as are homologous polypeptides and recombinant polypeptides (i.e., polypeptides having amino acid insertions, deletions, or substitutions) having greater than 40% sequence identity.

Subunits of CAP/HDAC complexes also can be identified by coimmunoprecipitation using antibodies against known CAP/HDAC subunits. Purification of CAP/HDAC subunits using coimmunoprecipitation has been described, for example, in: Jones P.L. et al. *Nature Genet* 19, 187-191 (1998); van der Vlag, J. and Otte, A.P. *Nature Genetics* 25, 474-478 (1999); Wade, P.A. et al. *Nature Genetics* 23, 62-66 (1999); Ng, H.H. et al. *Nature Genetics* 23, 58-61 (1999); and Spillane C. et al. *Curr Biol.* 10, 1535-1538 (2000).

Subunits of CAP/HDAC complexes also can be identified by yeast two-hybrid analyses using hybrid polypeptides containing known CAP/HDAC subunits. Use of the yeast two-hybrid system to identify CAP/HDAC subunits has been described, for example, in: Yadegari. R. et al. *Plant Cell* 12, 2367-2381 (2000); and Spillane C. et al. *Curr Biol.* 10, 1535-1538 (2000).

In some instances, suitable second polypeptide segments can be synthesized on the basis of consensus functional domains and/or conserved regions in polypeptides that are homologous subunits of a CAP/HDAC complex. Consensus domains and conserved regions can be identified by homologous polypeptide sequence analysis as described herein. The suitability of such synthetic polypeptides for use as a second polypeptide segment can be evaluated by the techniques described herein, or by evaluating the ability

of a synthetic polypeptide to effectively substitute for a corresponding subunit when expressed in a eukaryotic organism.

Many CAP/HDAC complexes and CAP/HDAC complex subunits are known to be conserved in plants, fungi and animals. Subunits of a CAP/HDAC complex in one organism can be used to identify homologous subunits in another organism, e.g., homologs of a subunit of a known CAP/HDAC complex can be identified by performing a BLAST query on a database of protein sequences. Those proteins in the database that have greater than 40% sequence identity are candidates for further evaluation for suitability as a second polypeptide segment. For example, the *Arabidopsis* polycomb group proteins FIE and MEA have significant sequence identity to the *Drosophila* proteins extra sex combs and enhancer of zeste. If desired, manual inspection of such candidates can be carried out in order to narrow the number of candidates for further evaluation. Manual inspection is carried out by selecting those candidates that appear to have domains suspected of being present in subunits of CAP/HDAC complexes.

Further evaluation can be carried out by creating a chimeric polypeptide having the candidate as the second segment, inserting the chimeric polypeptide into a eukaryotic organism, and evaluating the phenotypic effect of the chimeric polypeptide in the organism. If the desired phenotypic effect(s) is observed, the candidate is suitable as a second polypeptide segment.

A percent identity for any subject nucleic acid or amino acid sequence (e.g., any of the chimeric polypeptide first polypeptide segments, or second polypeptide segments described herein) relative to another "target" nucleic acid or amino acid sequence can be determined as follows. First, a target nucleic acid or amino acid sequence of the invention can be compared and aligned to a subject nucleic acid or amino acid sequence using the BLAST 2 Sequences (Bl2seq) program from the stand-alone version of BLASTZ containing BLASTN and BLASTP (e.g., version 2.0.14). The stand-alone version of BLASTZ can be obtained at <www.fr.com> or <www.ncbi.nlm.nih.gov>. Instructions explaining how to use BLASTZ, and specifically the Bl2seq program, can be found in the 'readme' file accompanying BLASTZ. The programs also are described in detail by Karlin et al. (*Proc. Natl. Acad. Sci. USA*, 87:2264 (1990) and 90:5873 (1993)) and Altschul et al. (*Nucl. Acids Res.*, 25:3389 (1997)).

Bl2seq performs a comparison between a subject sequence and a target sequence using either the BLASTN (used to compare nucleic acid sequences) or BLASTP (used to compare amino acid sequences) algorithm. Typically, the default parameters of a BLOSUM62 scoring matrix, gap existence cost of 11 and extension cost of 1, a word size of 3, an expect value of 10, a per residue cost of 1 and a lambda ratio of 0.85 are used when performing amino acid sequence alignments. The output file contains aligned regions of homology between the target sequence and the subject sequence. Once aligned, a length is determined by counting the number of consecutive nucleotides or amino acid residues (*i.e.*, excluding gaps) from the target sequence that align with sequence from the subject sequence starting with any matched position and ending with any other matched position. A matched position is any position where an identical nucleotide or amino acid residue is present in both the target and subject sequence. Gaps of one or more residues can be inserted into a target or subject sequence to maximize sequence alignments between structurally conserved domains.

The percent identity over a particular length is determined by counting the number of matched positions over that particular length, dividing that number by the length and multiplying the resulting value by 100. For example, if (*i*) a 500 amino acid target sequence is compared to a subject amino acid sequence, (*ii*) the Bl2seq program presents 200 amino acids from the target sequence aligned with a region of the subject sequence where the first and last amino acids of that 200 amino acid region are matches, and (*iii*) the number of matches over those 200 aligned amino acids is 180, then the 500 amino acid target sequence contains a length of 200 and a sequence identity over that length of 90% (*i.e.*, $180 \div 200 \times 100 = 90$). In some embodiments, the amino acid sequence of a second polypeptide segment has 40% sequence identity to the amino acid sequence of a subunit of a CAP/HDAC complex. In some embodiments, the amino acid sequence of a second polypeptide segment has greater than 40% sequence identity (e.g., > 80%, > 70%, > 60%, > 50% or > 40%) to the amino acid sequence of a subunit of a CAP/HDAC complex.

It will be appreciated that a nucleic acid or amino acid target sequence that aligns with a subject sequence can result in many different lengths with each length having its own percent identity. It is noted that the percent identity value can be rounded to the

nearest tenth. For example, 78.11, 78.12, 78.13, and 78.14 is rounded down to 78.1, while 78.15, 78.16, 78.17, 78.18, and 78.19 is rounded up to 78.2. It is also noted that the length value will always be an integer.

A partial list of nucleic acids encoding proteins that are subunits of CAP/HDAC complexes is shown in Table 1. The nucleic acids shown in Table 1 encode proteins that are subunits of CAP/HDAC complexes often referred to as polycomb group (PcG) complexes. Such proteins are candidates to be the second polypeptide segment.

Table 1. Polycomb Group Subunits

Genes	GI number-Source
Additional sex combs (Asx)	GI:3292939
Cramped	GI:5869804
Enhancer of Zeste (E(z))	GI:404864
Enhancer of polycomb	GI:3757890
Extra sex combs (Esc)	GI:2133657 or GI:1050997
At Epl1	GI:9989052
At Epl2	GI:3152596
ZmEpl01	GI:20152912
Multi sex combs (mxc)	GI:6746602
Pleiohomeotic (pho)	GI:3258627
Polycomb (Pc)	GI:129718
Polycomb-like (Pcl)	GI:521181
Polyhomeotic distal (mouse)	GI:1490546
Polyhomeotic proximal (php)	GI:730323
Posterior sexcombs (Psc)	GI:548613 or GI:103177
Sexcomb extra (Sce)	sequence unknown
Sex comb on midleg (Scm)	GI:1293574
Suppressor-2 of zeste	GI:236137 (partial)
Supressor of zeste 12 Su(z)12	GI:8131946
Su(z)2(D)	sequence unknown
Super sex combs (sxc)	sequence unknown
At Fis2	GI:4185501
At Emf2	GI:14276050
At Vrn2	GI:16945788
At MEA; At CLF; At E(Z)-likeA1; Mez1; Mez2; Mez3	GI:3089625
At Fie	GI:4567095
Zm Fie1	GI:18032004
Zm Fie2	GI:18032006

In some embodiments, a second polypeptide segment is the polypeptide encoded by the *Arabidopsis* Mea, FIS2, FIE, At E(Z)-likeA1, curly-leaf, or TSO1-like genes or homologs thereof. Polypeptides having modifications relative to these polypeptides also can be suitable second polypeptide segments.

Also useful are proteins that are subunits of SIN3/HDAC complexes, including, for example, Sin3, Rpd3 RbAp48, RbAp46, NcoR and SMRT. See e.g., Wolffe, A.P. et al., *Mol. Cell Biol.*, 19:5847-5860 (1999). A partial list of nucleic acids encoding proteins that are subunits of SIN3/HDAC complexes is shown in Table 2. Polypeptides having modifications relative to these polypeptides also are suitable second polypeptide segments.

Table 2. Subunits of Sin3/HDAC Complexes

Genes	GI number
Sin3	GI:9624449
STB1 (Sin3 binding protein)	GI:988311
STB2 (Sin3 binding protein)	GI:988309
Rpd3	GI:417699
SDS3 (suppressor of defective silencing 3)	GI:1480732
HD2A	GI:7489751
HD2B	GI:7716948
HDAC1	GI:2498443
HDAC2	GI:3023939
RbAp48	GI:3309245
RbAp46	GI:4506439
SMRT	GI:2136312
Tup1	GI:83454
Ume6	GI:6320413
N-CoR1 and 2 (nuclear receptor corepressor)	GI:5454138; GI:12643957
Ssn6	GI:117936
Mad1	GI:1708908
Mnt	GI:6754718
Mxi1	GI:1709194
Rox	GI:3914034
PSF (polypyrimidine tract-binding protein-associated splicing factor)	GI:10442545
NonO/p54(nrb)	GI:13124797
Ikaros	GI:3915731
Aiolos	GI:2150044
MBD1	GI:7305259
MBD2	GI:5929756
MBD3	GI:4505119
MBD4	GI:6754652
MeCP1 (PCM1)	GI:7710141
MeCP2	GI:1708973
Mi-2	GI:4557451
SAP18	GI:11433775; 5032067
SAP30	GI:11436724; 4506783
MTA-like	GI:6754644
KRAB-ZFP (Kruppel associated box)	GI:9625008

Also useful are proteins that are subunits of Mad-Max complexes, another group of CAP/HDAC complexes. Examples of Mad-Max complex subunits include Max-Mad-Mxi-Myc (basic HLH), mSin3a/B, HDAC1/2, N-CoR (nuclear receptor corepressor), and SMRT (silencing mediator of retinoic acid and thyroid hormone receptor). Also useful are proteins that are subunits of Tup1-Ssn6 complexes. Examples of Tup1-Ssn6 complex

subunits include Ume6, Tup1, Ssn6, Mig1, α2 or Crt1, and HDAC class I complexes (Rpd3, Hos1, Hos2). See e.g., Watson A.D. Genes & Dev., 14:2737-2744, (2000). Other suitable subunits can include Sin4, Srb8, Srb10, Srb11, and Med3. In other embodiments, polypeptides having modifications relative to the above polypeptides are suitable second polypeptide segments.

Exemplary nucleotide and/or amino acid sequences of CAP/HDAC subunit genes and/or polypeptides are shown in Table 7.

Arrangement of polypeptide segments

Segments of a chimeric HAT polypeptide are linked to one another by covalent bonds, typically peptide bonds. The segments can be linked directly, without any intervening amino acids between two segments. Alternatively, one segment can be linked indirectly to an adjacent segment by amino acid residues that are situated between the two adjacent segments and are themselves covalently linked to the adjacent segments. In some embodiments, there are one, two, three, four, five, six, seven, eight, nine or ten intervening amino acid residues. In other embodiments, there are fifteen, twenty, thirty, forty or fifty intervening residues. In some embodiments, an intervening segment can be a hinge domain. Typically, if there is an intervening segment, at least one of the amino acids in the intervening segment is a glycine. At least one glycine is preferred in order to promote structural flexibility of the spacer, and permit free rotation of the first polypeptide segment relative to the second polypeptide segment. An illustrative embodiment of an intervening segment is one having fifteen glycine residues positioned between the first polypeptide segment and the second polypeptide segment and covalently linked to each by a peptide bond.

An intervening peptide segment can be situated between the segments of a chimeric polypeptide of the invention in order to facilitate interaction between the histone in a nucleosome and the HAT of the chimeric polypeptide. Structural modeling can be used to predict whether an intervening peptide segment is useful in a chimeric HAT polypeptide. Structural modeling can be performed using software such as Rasmol 2.6, available from the UC Berkeley website http://mc2.CChem.Berkeley.EDU/Rasmol/v2.6/. For example, the theoretical distance between the first polypeptide segment of a chimeric

polypeptide and the surface of a nucleosome is modeled, based on the crystal structure of a nucleosome (histones H2A, H2B, H3 and H4, and a 147 nucleotide DNA), the crystal structure of the DNA binding domain of a TATA binding protein and the crystal structure of a *Tetrahymena* histone acetyltransferase GCN5 homologue, including the coenzyme Acetyl-CoA and the 11-mer N-terminal tail of histone H3. The TATA binding protein is modeled as it is situated on the DNA of the nucleosome. The HAT is modeled while adjacent to the tail of histone H3. Next, the distance from the closest surface of HAT to the nucleosome surface is calculated. Based on this example, an intervening peptide segment of at least 28 Å in length facilitates interaction between the HAT and histone yet maintains nucleosome interaction and histone modification. Twenty eight Å is approximately the same length as a peptide containing 15 amino acids. Structural flexibility of the intervening peptide segment can be enhanced by using at least one glycine amino acid and/or at least one alanine amino acid.

The first polypeptide segment of a chimeric polypeptide can be the N-terminal segment of a chimeric polypeptide of the invention. In such embodiments, the C-terminus of the first polypeptide segment can be covalently linked to the N-terminus of the second polypeptide segment, or can be covalently linked to the N-terminus of an intervening peptide segment, which can be schematically indicated at 1^{st}-2^{nd} or 1^{st}-i-2^{nd}, where "1^{st}" indicates the first polypeptide segment, "2^{nd}" indicates the second polypeptide segment and "i" indicates an optional intervening peptide segment.

In other embodiments, the first polypeptide segment can be the C-terminal segment of a chimeric polypeptide of the invention. In such embodiments, the C-terminus of the second polypeptide segment is covalently linked to the N-terminus of the first polypeptide segment, or can be covalently linked to the N-terminus of an intervening peptide segment, which can be schematically indicated as 2^{nd}-1^{st} or 2^{nd}-i-1^{st}.

A chimeric polypeptide of the invention optionally can possess additional amino acid residues at the amino-terminus or the carboxy-terminus. For example, 6x His-tag or FLAG® residues can be linked to a polypeptide at the amino-terminus. See e.g., U.S. Patent Nos. 4,851,341 and 5,001,912. As another example, a reporter polypeptide such as green fluorescent protein (GFP) can be fused to the carboxy-terminus of the chimeric polypeptide. See e.g., U.S. Patent No. 5,491,084.

With respect to polypeptides, "isolated" refers to a polypeptide that constitutes a major component in a mixture of components, e.g., 30% or more, 40% or more, 50% or more, 60% or more, 70% or more, 80% or more, 90% or more, or 95% or more by weight. Isolated polypeptides typically are obtained by purification from an organism that makes the polypeptide, although chemical synthesis is also feasible. Methods of polypeptide purification include, for example, chromatography or immunoaffinity techniques.

The amino acid sequence of either or both polypeptide segments of a chimeric HAT polypeptide can be a non-naturally occurring amino acid sequence. For example, the amino acid sequence of one polypeptide segment can be a naturally occurring sequence found in a particular species, while the amino acid sequence of the other polypeptide segment is a non-naturally occurring consensus amino acid sequence based on the naturally occurring sequences of homologs from different species.

A polypeptide of the invention can be detected by sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis followed by Coomassie Blue-staining or Western blot analysis using monoclonal or polyclonal antibodies that have binding affinity for the polypeptide to be detected.

Nucleic Acids Encoding a Chimeric Polypeptide

The present invention also includes nucleic acids encoding the above-described chimeric polypeptides. As used herein, nucleic acid refers to RNA or DNA, including cDNA, synthetic DNA or genomic DNA. The nucleic acids can be single- or double-stranded, and if single-stranded, can be either the coding or non-coding strand. As used herein with respect to nucleic acids, "isolated" refers to (i) a naturally-occurring nucleic acid encoding part or all of a polypeptide of the invention, but free of sequences, i.e., coding sequences, that normally flank one or both sides of the nucleic acid encoding polypeptide in a genome; (ii) a nucleic acid incorporated into a vector or into the genomic DNA of an organism such that the resulting molecule is not identical to any naturally-occurring vector or genomic DNA; or (iii) a cDNA, a genomic nucleic acid fragment, a fragment produced by polymerase chain reaction (PCR) or a restriction fragment.

Specifically excluded from this definition are nucleic acids present in mixtures of nucleic acid molecules or cells.

It should be appreciated that nucleic acids having a nucleotide sequence other than the specific nucleotide sequences disclosed herein can still encode a polypeptide having the exemplified amino acid sequence. The degeneracy of the genetic code is well known to those of ordinary skill in the art; i.e., for many amino acids, there is more than one nucleotide triplet that serves as the codon for the amino acid.

Nucleic acid constructs

Further provided are nucleic acid constructs comprising the above-described nucleic acid coding sequences. Such constructs can comprise a cloning vector. Cloning vectors suitable for use in the present invention are commercially available and are used routinely by those of ordinary skill in the art.

Nucleic acid constructs also can contain sequences encoding other polypeptides. Such polypeptides can, for example, facilitate the introduction or maintenance of the nucleic acid construct into a host organism. Other polypeptides also can affect the expression, activity, or biochemical or physiological effect of the encoded CBF polypeptide. Alternatively, other polypeptide coding sequences can be provided on separate nucleic acid constructs.

Nucleic acid constructs of the invention can comprise one or more regulatory elements operably linked to a nucleic acid coding sequence. Such regulatory elements can include promoter sequences, enhancer sequences, response elements or inducible elements that modulate expression of a nucleic acid sequence. As used herein, "operably linked" refers to positioning of a regulatory element in a construct relative to a nucleic acid coding sequence in such a way as to permit or facilitate expression of the encoded polypeptide. The choice of element(s) that can be included depends upon several factors, including, but not limited to, replication efficiency, selectability, inducibility, desired expression level, and cell or tissue specificity.

Suitable regulatory elements include promoters that initiate transcription only, or predominantly, in certain cell types. For example, promoters specific to vegetative tissues such as ground meristem, vascular bundle, cambium, phloem, cortex, shoot apical

meristem, lateral shoot meristem, root apical meristem, lateral root meristem, leaf primordium, leaf mesophyll, or leaf epidermis can be suitable regulatory elements. In other embodiments, a promoter specific to a reproductive tissue (e.g., fruit, ovule, seed, pollen, pistils, female gametophyte, egg cell, central cell, nucellus, suspensor, synergid cell, flowers, embryonic tissue, embryo, zygote, endosperm, integument, seed coat or pollen) is used. A cell type or tissue-specific promoter can drive expression of operably linked sequences in tissues other than the target tissue. Thus, as used herein a cell type or tissue-specific promoter is one that drives expression preferentially in the target tissue, but can also lead to some expression in other cell types or tissues as well. Methods for identifying and characterizing promoter regions in plant genomic DNA include, for example, those described in the following references: Jordano, et al., *Plant Cell*, 1:855-866 (1989); Bustos, et al., *Plant Cell*, 1:839-854 (1989); Green, et al., *EMBO J.*, 7:4035-4044 (1988); Meier, et al., *Plant Cell*, 3:309-316 (1991); and Zhang, et al., *Plant Physio.*, 110:1069-1079 (1996).

Exemplary reproductive tissue promoters include those derived from the following seed-genes: zygote and embryo LEC1 (see, Lotan (1998) *Cell* 93:1195-1205); suspensor G564 (see, Weterings. (2001) *Plant Cell* 13:2409-2425); maize MAC1 (see, Sheridan (1996) *Genetics*, 142:1009-1020); maize Cat3 (see, GenBank No. L05934; Abler (1993) *Plant Mol. Biol.*, 22:10131-1038); *Arabidopsis* viviparous-1 (see, Genbank No. U93215); *Arabidopsis* atmycl (see, Urao (1996) *Plant Mol. Biol.*, 32:571-57; Conceicao (1994) *Plant*, 5:493-505); and Brassica napus napin gene family, including napA (see, GenBank No. J02798; Josefsson (1987) JBL 26:12196-1301; Sjodahl (1995) *Planta*, 197:264-271). Other exemplary reproductive tissue-specific promoters include those derived from the pollen genes described in, for example: Guerrero (1990) *Mol. Gen. Genet.*, 224:161-168; Wakeley (1998) *Plant Mol. Biol.*, 37:187-192; Ficker (1998) *Mol. Gen. Genet.*, 257:132-142; Kulikauskas (1997) *Plant Mol. Biol.*, 34:809-814; and Treacy (1997) *Plant Mol. Biol.*, 34:603-611. Yet other suitable reproductive tissue promoters include those derived from the following embryo genes: Brassica napus 2s storage protein (see, Dasgupta (1993) *Gene*, 133:301-302); *Arabidopsis* 2s storage protein (see, GenBank No. AL161566); soybean β-conglycinin (see, GenBank No. S44893); Brassica napus oleosin 20kD gene (see, GenBank No. M63985); soybean oleosin A (see, Genbank

No. U09118); soybean oleosin B (see, GenBank No. U09119); soybean lectin1 (see, GenBank K00821); soybean Kunitz trypsin inhibitor 3 (see, GenBank No. AF233296); soybean glycinin1 (see, GenBank No. X15121); *Arabidopsis* oleosin (see, GenBank No. Z17657); maize oleosin 18kD (see, GenBank No. J05212; Lee (1994) Plant Mol. Biol. 26:1981-1987); and the gene encoding low molecular weight sulfur rich protein from soybean (see, Choi (1995) *Mol. Gen. Genet.*, 246:266-268). Yet other exemplary reproductive tissue promoters include those derived from the following endosperm genes: *Arabidopsis* Fie (see, GenBank No. AF129516); *Arabidopsis* Mea; *Arabidopsis* Fis2 (see, GenBank No. AF096096); rice Glu1 (see, GenBank No. M28156); and rice 26 kDa globulin (see, GenBank No. D50643). Yet other exemplary reproductive tissue promoters include those derived from the following genes: ovule BEL1 (see, Reiser (1995) *Cell*, 83:735-742; Ray (1994) *Proc. Natl. Acad. Sci. USA*, 91:5761-5765; GenBank No. U39944); central cell FIE (see, GenBank No. AF129516); flower primordia *Arabidopsis* APETALA1 (a.k.a. AP1) (see, Gustafson-Brown (1994) *Cell*, 76:131-143; Mandrel (1992) *Nature*, 360:273-277); flower *Arabidopsis* AP2 (see, Jofuku (1994) *Plant Cell* 6:1211-1225); *Arabidopsis* flower ufo, expressed at the junction between sepal and petal primordia (see, Bossinger (1996) *Development*, 122:1093-1102); fruit-specific tomato E8; a tomato gene expressed during fruit ripening, senescence and abscission of leaves and flowers (see, Blume (1997) *Plant J.*, 12:731-746); pistil-specific potato SK2 (see, Ficker (1997) *Plant Mol. Biol.*, 35:425-431); *Arabidopsis* DMC1 (see, GenBank No. U76670); and *Arabidopsis* DMT1 (see, Choi (2002) *Cell*, 109).

Suitable vegetative tissue promoters include those derived from the following genes: pea Blec4, active in epidermal tissue of vegetative and floral shoot apices of transgenic alfalfa; potato storage protein patatin gene (see, Kim (1994) *Plant Mol. Biol.*, 26:603-615; Martin (1997) *Plant J.*, 11:53-62); root Agrobacterium rhizogenes ORF13 (see, Hansen (1997) *Mol. Gen. Genet.*, 254:337-343); genes active during taro corm development (see, Bezerra (1995) *Plant Mol. Biol.*, 28:137-144); de Castro (1992) *Plant Cell*, 4:1549-1559); root meristem and immature central cylinder tobacco gene TobRB7 (see, Yamamoto (1991) *Plant Cell*, 3:371-382); ribulose biphosphate carboxylase genes RBCS1, RBCS2, and RBCS3A expressed in tomato leaves (see, Meier (1997) *FEBS Lett.*, 415:91-95); ribulose biphosphate carboxylase genes expressed in leaf blade and leaf

sheath mesophyll cells (see, Matsuoka (1994) *Plant J.*, 6:311-319); leaf chlorophyll a/b binding protein (see e.g., Shiina (1997) *Plant Physiol.*, 115:477-483; Casal (1998) *Plant Physiol.*, 116:1533-1538); *Arabidopsis* Atmyb5, expressed in developing leaf trichomes, stipules, in epidermal cells on the margins of young rosette and cauline leaves, and in immature seeds between fertilization and the 16 cell stage of embryo development and persists beyond the heart stage (see, Li (1996) *FEBS Lett.*, 379:117-121); a maize leaf-specific gene described by Busk (1997) *Plant J.*, 11:1285-1295; "SHOOTMERISTEMLESS" and "SCARECROW" genes active in developing shoot or root apical meristems (see e.g., Di Laurenzio (1996) *Cell*, 86:423-433; Long (1996) *Nature*, 379:66-69); 3-hydroxy-3-methylglutaryl coenzyme A reductase HMG2, expressed in meristematic tissue, and floral reductase HMG2, expressed in meristematic and floral (e.g., secretory zone of the stigma, mature pollen grains, gynoecium vascular tissue, and fertilized ovules) tissues (see, Enjuto (1995) *Plant Cell*, 7:517-527); meristem kn1-related genes from maize and other species (see, Granger (1996) *Plant Mol. Biol.*, 31:373-378; Kerstetter (1994) *Plant Cell*, 6:1877-1887; Hake (1995) *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 350:45-51; Lincoln (1994) *Plant Cell*, 6:1859-1876); and constitutive Cauliflower mosaic virus 35S.

Cell type or tissue-specific promoters derived from viruses also can be suitable regulatory elements. Exemplary viral promoters include: the tobamovirus subgenomic promoter (Kumagai (1995) *Proc. Natl. Acad. Sci. USA*, 92:1679-1683; the phloem-specific tungro bacilliform virus (RTBV) promoter; the cassava vein mosaic virus (CVMV) promoter, expressed most strongly in vascular elements, leaf mesophyll cells, and root tips (Verdaguer (1996) *Plant. Mol. Biol.*, 31:1129-1139).

In some embodiments, a nucleic acid construct of the invention contains a promoter and a recognition site for a transcriptional activator, both of which are operably linked to the coding sequence for a chimeric polypeptide. In these embodiments, transgenic organisms or mixtures of cells that express the chimeric polypeptide contain a second nucleic acid construct that encodes a transcriptional activator. A transcriptional activator is a polypeptide that binds to a recognition site on DNA, resulting in an increase in the level of transcription from a promoter associated in *cis* with the recognition site.

The recognition site for the transcriptional activator polypeptide is positioned with respect to the promoter so that upon binding of the transcriptional activator to the recognition site, the level of transcription from the promoter is increased. The position of the recognition site relative to the promoter can be varied for different transcriptional activators, in order to achieve the desired increase in the level of transcription.

Many transcriptional activators have discrete DNA binding and transcription activation domains. The DNA binding domain(s) and transcription activation domain(s) of transcriptional activators can be synthetic or can be derived from different sources (e.g., two-component system or chimeric transcriptional activators). In some embodiments, a two-component system transcriptional activator has a DNA binding domain derived from the yeast gal4 gene and a transcription activation domain derived from the VP16 gene of herpes simplex virus. In other embodiments, a two-component system transcriptional activator has a DNA binding domain derived from a yeast HAP1 gene and the transcription activation domain derived from VP16. Populations of transgenic organisms or cells having a first nucleic acid construct that encodes a chimeric polypeptide and a second nucleic acid construct that encodes a transcriptional activator polypeptide can be produced by transformation, transfection, or genetic crossing. See e.g., WO 97/31064.

A nucleic acid encoding a novel polypeptide of the invention can be obtained by, for example, DNA synthesis or the polymerase chain reaction (PCR). PCR refers to a procedure or technique in which target nucleic acids are amplified. PCR can be used to amplify specific sequences from DNA as well as RNA, including sequences from total genomic DNA or total cellular RNA. Various PCR methods are described, for example, in *PCR Primer: A Laboratory Manual*, Dieffenbach, C. & Dveksler, G., Eds., Cold Spring Harbor Laboratory Press, 1995. Generally, sequence information from the ends of the region of interest or beyond is employed to design oligonucleotide primers that are identical or similar in sequence to opposite strands of the template to be amplified. Various PCR strategies are available by which site-specific nucleotide sequence modifications can be introduced into a template nucleic acid.

Nucleic acids of the present invention can be detected by methods such as ethidium bromide staining of agarose gels, Southern or Northern blot hybridization, PCR

or *in situ* hybridizations. Hybridization typically involves Southern or Northern blotting (see, for example, sections 9.37-9.52 of Sambrook et al., 1989, *"Molecular Cloning, A Laboratory Manual"*, 2nd Edition, Cold Spring Harbor Press, Plainview; NY). Probes should hybridize under high stringency conditions to a nucleic acid or the complement thereof. High stringency conditions can include the use of low ionic strength and high temperature washes, for example 0.015 M NaCl/0.0015 M sodium citrate (0.1X SSC), 0.1% sodium dodecyl sulfate (SDS) at 65°C. In addition, denaturing agents, such as formamide, can be employed during high stringency hybridization, e.g., 50% formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 with 750 mM NaCl, 75 mM sodium citrate at 42°C.

Eukaryotic Organisms

The term "host" or "host cell" includes not only prokaryotes, such as *E. coli*, but also eukaryotes, such as fungal, insect, plant and animal cells. Animal cells include, for example, COS cells and HeLa cells. Fungal cells include yeast cells, such as *Saccharomyces cereviseae* cells. A host cell can be transformed or transfected with a DNA molecule (e.g., a vector) using techniques known to those of ordinary skill in this art, such as calcium phosphate or lithium acetate precipitation, electroporation, lipofection and particle bombardment. Host cells containing a vector of the present invention can be used for such purposes as propagating the vector, producing a nucleic acid (e.g., DNA, RNA, antisense RNA) or expressing a polypeptide or fragments thereof.

Plants

Among the eukaryotic organisms featured in the invention are plants containing an exogenous nucleic acid that encodes a polypeptide of the invention, e.g., nucleic acids encoding a polypeptide having an amino acid sequence as shown in Table 9 or in Table 11.

Accordingly, a method according to the invention comprises introducing a nucleic acid construct as described herein into a plant. Techniques for introducing exogenous nucleic acids into monocotyledonous and dicotyledonous plants are known in the art, and include, without limitation, *Agrobacterium*-mediated transformation, viral vector-

mediated transformation, electroporation and particle gun transformation, e.g., U.S. Patents 5,204,253 and 6,013,863. If a cell or tissue culture is used as the recipient tissue for transformation, plants can be regenerated from transformed cultures by techniques known to those skilled in the art. Transgenic plants can be entered into a breeding program, e.g., to introduce a nucleic acid encoding a polypeptide into other lines, to transfer the nucleic acid to other species or for further selection of other desirable traits. Alternatively, transgenic plants can be propagated vegetatively for those species amenable to such techniques. Progeny includes descendants of a particular plant or plant line. Progeny of an instant plant include seeds formed on F_1, F_2, F_3, and subsequent generation plants, or seeds formed on BC_1, BC_2, BC_3, and subsequent generation plants. Seeds produced by a transgenic plant can be grown and then selfed (or outcrossed and selfed) to obtain seeds homozygous for the nucleic acid encoding a novel polypeptide.

A suitable group of plants with which to practice the invention include dicots, such as safflower, alfalfa, soybean, rapeseed (high erucic acid and canola), or sunflower. Also suitable are monocots such as corn, wheat, rye, barley, oat, rice, millet, amaranth or sorghum. Also suitable are vegetable crops or root crops such as potato, broccoli, peas, sweet corn, popcorn, tomato, beans (including kidney beans, lima beans, dry beans, green beans) and the like. Thus, the invention has use over a broad range of plants, including species from the genera *Anacardium*, *Arachis*, *Asparagus*, *Atropa*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Carthamus*, *Cocos*, *Coffea*, *Cucumis*, *Cucurbita*, *Daucus*, *Elaeis*, *Fragaria*, *Glycine*, *Gossypium*, *Helianthus*, *Heterocallis*, *Hordeum*, *Hyoscyamus*, *Lactuca*, *Linum*, *Lolium*, *Lupinus*, *Lycopersicon*, *Malus*, *Manihot*, *Majorana*, *Medicago*, *Nicotiana*, *Olea*, *Oryza*, *Panicum*, *Pannesetum*, *Persea*, *Phaseolus*, *Pistachia*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Ricinus*, *Secale*, *Senecio*, *Sinapis*, *Solanum*, *Sorghum*, *Theobromus*, *Trigonella*, *Triticum*, *Vicia*, *Vitis*, *Vigna* and *Zea*.

Chimeric polypeptides of the invention can be expressed in plants in a cell- or tissue-specific manner according to the regulatory elements chosen to include in a particular nucleic acid construct present in the plant. Suitable cells, tissues and organs in which to express a chimeric polypeptide of the invention include, without limitation, egg cell, central cell, synergid cell, zygote, ovule primordia, nucellus, integuments, endothelium, female gametophyte cells, embryo, axis, cotyledons, suspensor, endosperm,

seed coat, ground meristem, vascular bundle, cambium, phloem, cortex, shoot or root apical meristems, lateral shoot or root meristems, floral meristem, leaf primordia, leaf mesophyll cells, and leaf epidermal cells, e.g., epidermal cells involved in forming the cuticular layer.

Fungi

Other eukaryotic organisms featured in the invention are fungi containing an exogenous nucleic acid that encodes a chimeric polypeptide of the invention, e.g., nucleic acids encoding a polypeptide having the amino acid sequence as shown in Table 9 or in Table 11.

Accordingly, a method according to the invention comprises introducing a nucleic acid construct as described herein into a fungus. Techniques for introducing exogenous nucleic acids into many fungi are known in the art, e.g., U.S. Patents 5,252,726 and 5,070,020. Transformed fungi can be cultured by techniques known to those skilled in the art. Such fungi can be used to introduce a nucleic acid encoding a polypeptide into other fungal strains, to transfer the nucleic acid to other species or for further selection of other desirable traits.

A suitable group of fungi with which to practice the invention include fission yeast and budding yeast, such as *Saccharomyces cereviseae*, *S. pombe*, *S. carlsbergeris* and *Candida albicans*. Filamentous fungi such as *Aspergillus spp.* and *Penicillium spp.* also are useful.

Animals

Other eukaryotic organisms featured in the invention are animals (e.g., insects such mosquitoes and flies; fish; and non-human mammals such as rodents, bovines and porcines) that contain an exogenous nucleic acid that encodes a chimeric polypeptide of the invention, e.g., nucleic acids encoding a polypeptide having the amino acid sequence as shown in Table 9 or in Table 11. A variety of techniques known in the art can be used to generate such transgenic animals. Such techniques typically involve generating a plurality of animals whose genomes can be screened for the presence or absence of the transgene. For example, a transgene can be introduced into a non-human mammal by

pronuclear microinjection (U.S. Patent No. 4,873,191), retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci. USA, 82:6148, 1985), gene targeting into embryonic stem cells (Thompson et al., Cell, 56:313, 1989), electroporation of embryos (Lo, Mol. Cell. Biol., 3:1803, 1983), and transformation of somatic cells *in vitro* followed by nuclear transplantation (Wilmut et al., Nature, 385(6619):810-813, 1997; and Wakayama et al., Nature, 394:369-374, 1998). When using mice to make a transgenic animal, suitable genetic backgrounds for use in making founder lines include, without limitation, C57B6, SJL/J, FVB/N, 129SV, BALB/C, C3H, and hybrids thereof.

Methods of gene profiling

In another aspect, the invention provides a method in which gene function can be determined from changes in an organism's gene expression profile. The method involves expressing a chimeric polypeptide in a specific cell type, tissue or organ in an organism or population of organisms. The organism can be, for example, an animal, plant, or fungus. The term "specific cell type" refers to cells that have one or more characteristics that distinguish them from the other cells in an organism, or from other cells in a mixture of cells. Distinguishing features can include, for example, physical location, cell division rate, developmental stage, differentiation status, macromolecular composition, gene expression profile, protein expression profile, particular cell type, or presence or absence of a particular polypeptide. Specific cell types can be found in an organ, tissue, or tissue or cell culture, e.g., egg cells from embryo sacs, scutellar cells of a mature kernel, cells containing seed storage proteins from cotyledons and rapidly dividing fibroblasts from skin. Specific cell types also can be found in more than one organ, tissue, or tissue or cell culture, e.g., meristematic cells from plant shoot and root apices, and mucosal cells from the large intestine and the nasal cavity.

The method typically involves introducing an exogenous nucleic acid encoding the chimeric polypeptide into an organism. In some embodiments, the exogenous nucleic acid contains a regulatory element that directs expression of the chimeric polypeptide in specific cell types. In other embodiments, the exogenous nucleic acid is situated in the genome of the target organism such that expression of the chimeric polypeptide is

governed by native transcriptional regulatory elements (e.g., a native cell type-specific promoter).

In yet other embodiments, the nucleic acid construct encoding a chimeric polypeptide contains a recognition site for a transcriptional activator. In these embodiments, transgenic organisms or mixtures of cells that express the chimeric polypeptide contain a second nucleic acid construct that encodes the transcriptional activator, and one or more regulatory elements that facilitate expression of the transcription activator in a specific cell type. Thus, in these embodiments, the exogenous transcription activator is expressed in specific cells, and in turn activates transcription of the chimeric polypeptide in those cells. Populations of transgenic organisms or cells having a first nucleic acid construct that encodes a chimeric polypeptide and a second nucleic acid construct that encodes a transcriptional activator can be produced by transformation, transfection, or genetic crossing.

Cell type-specific expression of a chimeric polypeptide can alter an organism's gene expression profile (i.e., the cell types in which particular sets of genes are transcribed, and the level at which such genes are transcribed) relative to organisms that do not express the chimeric polypeptide. Alterations in gene expression profile can be manifested in changes in the macromolecular (e.g., RNA, protein, chemical) composition of organisms that express a chimeric polypeptide in a cell-specific manner. The skilled artisan can measure the RNA or protein composition of specific cells using routine techniques such as, for example, thin layer or gas-liquid chromatography, gel electrophoresis of protein extracted from appropriate cells, and gel electrophoresis of RNA extracted from appropriate cells. The skilled artisan can measure the expression of particular genes or proteins using the above-mentioned methods alone or in combination with, for example, protein immunochemistry or nucleic acid hybridization assays using electrophoretically or chromatographically separated macromolecules, microarray analysis, or specific RT-PCR. The above-described techniques can provide quantitative, semi-quantitative or qualitative detection of gene expression. Alterations in gene expression profile can be detected by comparing the gene expression profiles of, for example, a transgenic organism that expresses the chimeric polypeptide in specific cells

and an organism that lacks the nucleic acid construct or does not express the chimeric polypeptide.

Once the transcriptional and/or translational activity of a set of genes has been determined in a specific cell type and/or at a desired time, the function of the set of genes can be assigned to particular developmental, physiological and/or biochemical pathways. In addition, a microarray containing the set of genes, or a subset thereof, can be made. See e.g., U.S. patents 5,424,186 and 6,156,501. The microarray can contain a plurality of oligonucleotides, each oligonucleotide representing a portion of the sequence of one gene from the set of genes. Each of the oligonucleotides is coupled to a solid substrate at a known location. The substrate can be silica, polymeric materials, glass, beads, slides or chips. Such microarrays can be used, for example, to determine the level of transcription of the set of genes in other cell types and thereby identify genes whose transcription is repressed solely in the specific cell type. Such genes are suitable targets for further manipulation. For example, genes that are inactivated solely during fruit maturation can be targeted for a modification that results in continued expression of such genes for an additional period of time, in order to delay fruit ripening and/or increase fruit size.

Methods for modulating gene expression

In another aspect, the invention provides methods for modulating gene expression in an organism. Modulating gene expression involves expressing a chimeric polypeptide in specific cells in an organism or population of organisms. The organism can be, for example, yeast or a plant.

An exogenous nucleic acid encoding a chimeric polypeptide is introduced into an organism. In some embodiments, the exogenous nucleic acid contains a regulatory element that directs expression of the chimeric polypeptide in specific cells or tissues. In other embodiments, the exogenous nucleic acid is situated in genome of the target organism such that expression of the chimeric polypeptide is governed by native transcriptional regulatory elements (e.g., a native cell type or tissue-specific promoter).

In yet other embodiments, the nucleic acid construct that encodes a chimeric polypeptide contains a recognition site for a transcriptional activator. In these embodiments, transgenic organisms or mixtures of cells that express the chimeric

polypeptide contain a second nucleic acid construct that encodes a transcriptional activator. The second nucleic acid construct contains a regulatory element that directs expression of the transcription activator in specific cells. Thus, in these embodiments, the exogenous transcription activator is expressed in specific cells or tissues, and in turn activates transcription of the chimeric polypeptide in those cells. Populations of transgenic organisms or cells having a nucleic acid construct that encodes a chimeric polypeptide and a nucleic acid construct that encodes a transcriptional activator polypeptide can be produced by transformation, transfection, or genetic crossing.

By expressing a chimeric polypeptide in specific cells, it is possible to modulate gene expression in an organism (e.g., by derepressing genes that normally are transcriptionally inactive). An organism or cell exhibiting modulated gene expression can have compositional (e.g., protein, nucleic acid, lipid, saccharide), developmental and phenotypic alterations relative to organisms or cells that do not express the chimeric polypeptide. For example, modulated gene expression in plants can alter seed development, seed yield, seed composition, endosperm development, embryo development, cotyledon development, seed size, flowering time, plant size, leaf size, leaf shape, plant fertility, apical dominance, floral organ identity, root development, or organ composition. In plants, cell type-specific expression of chimeric polypeptides also can cause fertilization independent endosperm development and fertilization independent seed development.

In some embodiments, seed development can be altered by expressing a chimeric polypeptide in the developing ovule or seed of a plant. In such embodiments, the chimeric polypeptide can modulate endosperm and/or embryo development; developing seed in such plants can exhibit altered endosperm and/or altered embryo development; and plants can exhibit altered seed yield (by number and / or mass). The effects of expressing a chimeric polypeptide on seed development can be enhanced when DNA methylation is reduced. DNA methylation can be reduced, e.g., by mutation of or antisense nucleic acid interference with a gene encoding a DNA methyltransferase. Exemplary plant DNA methyltransferase genes include Met1, Cmt3, Zmet2, Drm1, Drm2 (Vielle-Calzada et al. (1999) *Genes & Dev.* 13:2971-2982; Richards et al. (2000) US patent 6,153,741; Dellaporta and Chen (2000) US Patent 6,011,200; Vinkenoog et al.

(2000) *The Plant Cell* 12:2271-2282; Luo et al. (2000) *Proc. Natl. Acad. Sci. USA* 97:10637-10642; Jackson et al. (2002) *Nature* 416:556-560). DNA methylation also can be reduced by mutation of or antisense nucleic acid interference with certain genes that encode chromatin associated proteins that have a role in DNA methylation. Such genes include Ddm1 (see Jeddeloh et al. (1999) *Nature Genetics* 22:94-97) and Kyp (see Jackson et al. (2002) *Nature* 416:556-560). In these embodiments, plants can have altered seed yield by mass. Mutations of or antisense nucleic acid interference with other genes, such as Mom (see Amedeo et al. (2000) *Nature* 405:203-206), that have a post-DNA methylation role in DNA methylation state also can enhance the effects of expressing a chimeric polypeptide on seed development.

In some embodiments, the exogenous nucleic acid contains a regulatory element that directs expression of the chimeric polypeptide to specific cells or tissues.

In yet other embodiments, the nucleic acid construct that encodes a chimeric polypeptide contains a recognition site for a transcriptional activator. In these embodiments, transgenic organisms or mixtures of cells that express the chimieric polypeptide contain a second nucleic acid construct that encodes a transcriptional activator. The second nucleic acid construct contains a regulatory element that directs expression of the transcription activator in specific cells. Thus, in these embodiments, the exogenous transcription activator is expressed in specific cells or tissues, and in turn activates transcription of the chimeric polypeptide in those cells. Populations of transgenic organisms or cells having a nucleic acid construct that encodes a chimeric polypeptide and a nucleic acid construct that encodes a transcriptional activator polypeptide can be produced by transformation, transfection, or genetic crossing.

Methods of making sterile plants

In another aspect, the invention provides methods for making sterile plants by introducing an exogenous nucleic acid encoding a chimeric polypeptide. In some embodiments, the exogenous nucleic acid contains a regulatory element that directs expression of the chimeric polypeptide in reproductive cells. In other embodiments, the exogenous nucleic acid is situated in genome of the target organism such that expression

of the chimeric polypeptide is governed by a native transcriptional regulatory element that facilitates transcription in reproductive cells.

In yet other embodiments, the nucleic acid construct that encodes a chimeric polypeptide contains a recognition site for a transcriptional activator. In these embodiments, transgenic plants that express the chimeric polypeptide contain a second nucleic acid construct that encodes a transcriptional activator and one or more regulatory elements that facilitate expression of the transcription activator in plant reproductive cells. Thus, in these embodiments, the transcription activator is expressed in plant reproductive cells, which in turn activates transcription of the chimeric polypeptide in reproductive cells. Transformation and/or genetic crosses, for example, can produce plants that contain a nucleic acid construct that encodes a chimeric polypeptide and a nucleic acid construct that encodes a transcriptional activator polypeptide. Expressing a chimeric polypeptide in plant reproductive cells can affect the affect the reproductive and / or developmental processes and prevent the production of viable embryos from female reproductive tissues.

The invention is further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES

Example 1: Polypeptides having histone acetyltransferase activity.

Polypeptides are tested for histone acetyltransferase activity using assays previously described (see Brownell, J. and Allis, C. *Proc. Natl. Acad. Sci. USA*, 92:6364-6368 (1995); Brownell, J. E. et al. *Cell*, 84:843-851 (1996)). Coding sequences of candidate polypeptides are cloned into an appropriate expression vector, the expression vector is introduced into a bacterial host strain, expression of the gene is induced and protein extract is prepared. The extracts are incubated with calf thymus histones and [^{3}H]-acetyl-Coenzyme A. Radioactivity transferred to the histone substrate in an extract-dependent manner is quantified by liquid scintillation counting. Candidate polypeptides that transfer radioactivity to the histone substrate compared to positive controls (extracts from hosts expressing known HAT polypeptides) and negative controls (extract alone, histones without extract and

comparable vector-only) have HAT activity. Alternatively, plant HAT activity is tested by determining whether expression of the corresponding cDNA is sufficient to rescue a yeast HAT mutant.

Example 2: Polypeptides having histone deacetylase activity.

Polypeptides are tested for histone deacetylase activity using assays previously described by van der Vlag, J. and Otte A.P. in *Nature Genetics*, 25:474-478 (1999). Coding sequences of candidate polypeptides are cloned into an appropriate expression vector, the expression vector is introduced into a bacterial host strain, expression of the gene is induced and protein extract is prepared. The extracts are incubated with [^{3}H]-acetylated histones or histone segments for 3-6 hours at 37 °C under shaking conditions in a buffer containing 20 mM Tris.-HCl, pH 7.4, ad 50 mM NaCl. The reaction is stopped by adding 7.7 mM HCl/1.2M acetic acid, and extracted with ethyl acetate. After centrifugation, the ethyl acetate fraction is counted in a liquid scintillation counter. Candidate polypeptides that remove radioactivity from the histone substrate compared to positive controls (extracts from hosts expressing known HDAC polypeptides) and negative controls (extract alone, histones without extract, vector-only, and parallel trichostatin A-containing reactions) have HDAC activity. Alternatively, HDAC activity is tested by determining whether expressing a candidate HDAC polypeptide (e.g., using an nucleic acid construct containing the corresponding cDNA clone) in a yeast HDAC mutant can rescue the mutant phenotype.

Example 3: Chimeric HAT Nucleic Acid construct pFIE-15G-ESA1.

The chimeric HAT gene construct was constructed using standard molecular biology techniques. The construct contains the coding sequence for the *Arabidopsis* FIE polypeptide and the coding sequence for a truncated *Arabidopsis* HAT polypeptide linked in frame by a DNA fragment encoding fifteen glycine residues. The FIE coding sequence was obtained from plasmid pFIE3.6. The *Arabidopsis* FIE polypeptide is a homolog of the Drosophila polycomb protein extra sex combs (esc) (see Ohad et al., *Plant Cell*, 11:407-415 (1999)). The *Arabidopsis* HAT polypeptide

AtESA1 is a homolog of the yeast ESA1 polypeptide. pFIE-15G-ESA1 contains 5 binding sites for the DNA binding domain of the Gal4 transcription factor (UAS_{GAL4}) located 5' to a CaMV35S minimal promoter. The CaMV35S minimal promoter is located 5' to the FIE coding sequence. A DNA fragment encoding fifteen glycine residues is present, in frame, at the 3' end of the UAS-FIE DNA sequence, followed, in frame, by a DNA fragment encoding an *Arabidopsis* homologue of ESA1.

The coding sequence for a truncated AtESA1 was fused to the 3'-end of the FIE coding sequence by fusion PCR (Levin HL, *Mol. Cell Biol.*, 15:3310-3317 (1995)). Two intermediate PCR products were generated for this purpose. The first intermediate product contained a coding sequence for FIE having a 15-glycine spacer fused to its carboxy-terminus. This product was generated using two synthetic oligonucleotides and a pFIE3.6 DNA template. Similarly, the second intermediate PCR product contained a coding sequence for AtESA1 having a 15 glycine spacer fused to its amino-terminus. This product was generated using two synthetic oligonucleotides and a pAtESA1 cDNA template. The two intermediate products were then fused to each other in a final round of PCR using a set of synthetic primers that introduced a BglII site at the 5' end of the fusion and a XhoI site at the 3' end of the fusion. The resultant PCR product contained a chimeric sequence encoding a fusion peptide in which the amino-terminus of the FIE coding sequence is linked by a 15-glycine spacer to the carboxy-terminus the truncated AtESA1 coding sequence. This final PCR product was digested with BglII and XhoI and cloned into the Ti-plasmid vector pCRS304-5UAS which was previously digested with BamHI and XhoI. The resulting plasmid was named pCRS304-5UAS-FIE-15G-ESA1. The transgene was designated FIE-15G-ESA1. The amino acid sequence of the chimeric polypeptide encoded by the transgene is shown in Table 9 and the nucleotide sequence of the transgene is shown in Table 10.

Thus, pCRS304-5UAS-FIE-15G-ESA1 encodes a chimeric polypeptide having an *Arabidopsis thaliana* FIE polypeptide and a truncated *Arabidopsis thaliana* HAT polypeptide, linked by an intervening peptide spacer of 15 glycine residues. The plasmid contains 5 copies of the Gal4 upstream activator sequence (UAS_{GAL4}) located 5' and operably linked to the CaMV35S minimal promoter. This

in turn is located 5' and operably linked to the FIE-15G-ESA1 coding sequence. The binding of a transcription factor that possesses a Gal4 DNA binding domain to the Gal4 UAS is necessary for transcriptional activation.

Example 4: Chimeric HAT Nucleic Acid construct pMEA-15G-ESA1.

The chimeric HAT gene construct pMEA-15G-ESA1 was constructed using standard molecular biology techniques. The construct contains the coding sequence for the *Arabidopsis* MEA polypeptide and the coding sequence for an *Arabidopsis* HAT polypeptide joined in frame by a DNA fragment encoding fifteen glycine residues. The MEA coding sequence was obtained from plasmid pCB1(MEA-cDNA) (Kiyosue, T., et al. (1999) *Proc. Natl. Acad. Sci. USA* 96:4186-4191). The *Arabidopsis* MEA polypeptide is a homolog of the Drosophila polycomb protein Enhancer of zeste (E(z)) (see Grossniklaus, U., et al. (1998) *Science* 280: 446-450.; Kiyosue, T., et al. (1999) *Proc. Natl. Acad. Sci. USA* 96:4186-4191). The *Arabidopsis* HAT polypeptide AtESA1 is a homolog of the yeast ESA1 polypeptide. The pMEA-15G-ESA1 plasmid contains 5 binding sites for the DNA binding domain of the Gal4 transcription factor (UAS_{GAL4}) located 5' to a CaMV35S minimal promoter. The CaMV35S minimal promoter is located 5' to the MEA coding sequence. A DNA fragment encoding fifteen glycine residues is present, in frame, at the 3' end of the UAS- MEA DNA sequence, followed, in frame, by a DNA fragment encoding an *Arabidopsis* homologue of ESA1.

The AtESA1 coding sequence was fused to the 3'-end of the MEA coding sequence by standard cloning techniques. Two intermediate PCR products were generated for this purpose. The first intermediate product contained the MEA coding sequence, flanked on either side by a BamHI restriction site. The BamHI sites were generated by incorporation into the PCR primer sequences. The first intermediate PCR product was digested with BamHI restriction enzyme and was cloned into the T-DNA expression vector pCRS304-5USAL at its unique BamHI site. The resultant plasmid was named pCRS304-5USAL-MEA-no 3'UTR.

The second intermediate PCR product contained a coding sequence for AtESA1 having a 15-glycine spacer fused to its amino terminus. The second intermediate PCR product was generated using two synthetic oligonucleotides and

the pAtESA1-cDNA template. The second PCR product was flanked by a unique SmaI site at its 5' end and by a unique XhoI site at its 3'end. These cloning sites were generated by incorporation into the PCR primer sequences. The second PCR product was digested with SmaI and XhoI, and was cloned into the plasmid pCRS304-5USAL-MEA-no 3'UTR between the unique restriction sites SmaI and XhoI. The resultant plasmid was named pCRS304-5USAL-MEA-ESA1. The transgene was designated MEA-15G-ESA1. The amino acid sequence of the chimeric polypeptide encoded by the transgene is shown in Table 11 and the nucleotide sequence of the transgene is shown in Table 12.

Thus, pCRS304-5UAS-MEA-15G-ESA1 encodes a chimeric polypeptide having an *Arabidopsis thaliana* MEA polypeptide and an *Arabidopsis thaliana* HAT polypeptide, joined by an intervening peptide spacer of 15 glycine residues. The plasmid contains 5 copies of the Gal4 upstream activator sequence (UAS_{GAL4}) located 5' and operably linked to the CaMV35S minimal promoter. This in turn is located 5' and operably linked to the MEA-15G-ESA1 coding sequence. The binding of a transcription factor that possesses a Gal4 DNA binding domain to the Gal4 UAS is necessary for transcriptional activation.

Example 5: Transgenic plants.

The pCRS304-5UAS-FIE-15G-ESA1 plasmid and the pCRS304-5UAS-MEA-15G-ESA1 were independently introduced into *Arabidopsis* WS by *Agrobacterium tumefaciens* mediated transformation using the floral infiltration technique essentially as described in Bechtold, N. et al., *C.R. Acad. Sci. Paris*, 316:1194-1199 (1993). Several transformed plants, designated FE #1, FE #2, and ME #1, were selected for further study. The FIE-15G-ESA1 gene and the MEA-15G-ESA1 gene were then transcriptionally activated in specific target cells and tissues by crossing with two-component enhancer trap lines expressing a chimeric Gal4-VP16 activator protein (Haseloff et al.). In each activator line there is also a UAS_{GAL4}-GFP (green fluorescent protein) reporter gene.

Example 6: Two-component activation lines.

The two-component system for activating target gene expression was first utilized in Drosophila and subsequently adopted for use in plants (see Bennett et al. (1998) US Patent No. 5,801,027; Liu et al. (1999) US Patent No. 5,968,793); Bennett et al. (2000) US Patent No. 6,127,606; Haseloff and Hodge (2001) US Patent No. 6,255,558). The two-component system typically consists of two independent transcription units: an activator gene and a target gene. The activator gene encodes a transcriptional activator, a DNA binding protein gene such as Gal4-VP16, operably linked to a plant or animal promoter. The target gene has a protein coding sequence, such as a cDNA, operably linked to a promoter that has multiple copies of an upstream activator sequence element (UAS_{GAL4}) to which the transcriptional activator protein can bind. A target gene can be activated genetically by crossing a target gene-containing plant with an activation gene-containing plant (i.e., from an "activator line"). Alternatively, a target gene in a cell, tissue, or whole organism can be activated by transforming with an activation gene containing vector.

An extensive collection of *Arabidopsis* two-component activation lines has been produced and described by Dr. Jim Haseloff et al. (see http://www.plantsci.cam.ac.uk/Haseloff/IndexCatalogue.html), and individual lines are available from the *Arabidopsis* Biological Resource Center (see http://www.*Arabidopsis*.org/abrc/haseloff.htm). The activator lines were produced using a T-DNA based enhancer trap strategy. In this system the Gal4-VP16 gene containing a CaMV35S minimal promoter can be transcriptionally activated when T-DNA is inserted proximal to an endogenous enhancer element. Enhancer activity is revealed by the trans-activation of a UAS_{GAL4}-GFP reporter gene. Each activation line in the Haseloff collection contains one or more random T-DNA insertions in the *Arabidopsis* genome resulting in cell, tissue, or organ specific expression of a UAS_{GAL4}-GFP reporter gene. The amino acid sequence of the $GAL4_{UAS}$-VP16 activator protein is shown in Table 8.

Six publicly available *Arabidopsis* two-component activation lines are described in Table 3 including J2592, J0661, Q2500, M0164, J2301 and J2921.

Table 3. Activation line GFP expression pattern

Haseloff activation line reference number and ABRC seed stock number	Ovule and seed development	Root	Other	GFP Intensity
J2592 (CS9180)	Prefertilization ovule: ovule, funiculus and placenta. Developing seed: developing embryo and mature embryo.	Root cap, root epidermal cells.	Seedling: shoot and root epidermis, root cortex and root cap; hypocotyl, petiole epidermis, expanded cotyledon and leaf vasculature; stem epidermis and rosette leaf vasculature. Flower: sepal, petal and ovary vasculature; epidermis of mature sepal, petal, filament and ovary; stigma.	Medium
M0164 (CS9307)	Mature embryo.	Root: weak patchy expression in vasculature of primary root.	Seedling: strong in shoot apex, rosette leaf and petiole vasculature. Weak in cotyledon vasculature. Silique: older siliques only.	High
Q2500 (CS9135)	Ovule: prefertilization ovule. Seed: chalazal end of developing seed, seed coat and young embryo.	Root: vasculature.	Seedling: vasculature of hypocotyl, expanded cotyledons and first leaves. Flower: petal vasculature, placenta.	High

J0661 (CS9141)	Developing seed: funiculus, embryo.	Root: root vasculature.	Seedling: vasculature including root, hypocotyl, expanded cotyledons, rosette leaf vasculature, petiole. Cauline leaf vasculature. Flower: floral organ vasculature including pedicel, sepal, petal, filament and pistil.	Medium
J2921 (CS9194)		Root: weak patchy expression in root; weak in root hair; strong in root vasculature and root tip; strong in junctions where lateral roots form.	Flower: broad expression in epidermis of immature buds; GFP decreases and becomes restricted to the ovary as the flower matures; weak expression in sepal and petal vasculature.	Medium
J2301 (CS9173)	Seed: seedcoat; GFP increases as silique matures; GFP detectable at suspensor end of embryo.	Root: very strong in root tip; weak in root cortex; root epidermis.	Seedling: weak throughout seedling vasculature; strong in leaf trichomes; also detected in atrichoblasts. Flower: base of sepal and petal, ovary epidermis, style.	Medium

Each activation line displays a characteristic pattern of GFP accumulation in seedlings, vegetative organs and reproductive organs. GFP images are publicly available at http://www/plantsci.cam.ac.uk./Haseloff/GAL4 and were independently

confirmed. For example, in line J2592 GFP expression was detectable in young seedlings in the shoot and root epidermis, root cortex and root cap but not in the root apical meristem. GFP was also observed in seedling hypocotyl, petiole epidermis, expanded cotyledon and leaf vasculature. Low intensity GFP was detectable in the stem epidermis as well as in rosette leaf vasculature. GFP was observed in J2592 flowers including the vasculature of the sepal, petal and ovary and in the epidermis of the mature sepal, petal, filament, ovary and in stigmatic papillae. A low level of GFP was detected in the pedicel. GFP was observed in pre-fertilization ovules and in the funiculus and placenta. In fertilized seed GFP was detectable in developing seeds and in mature embryos. GFP expression patterns were observed to vary in some progeny of J2592.

In line M0164 seedlings, GFP expression was observed in the vasculature of the primary root. No expression was detectable in the root cap. Relatively intense GFP expression was observed in the shoot apex and in leaf and petiole vasculature. Low intensity GFP expression was observed in the cotyledon vasculature. In developing seed GFP was detectable in embryos. GFP expression was not detectable in the seed coat or endosperm.

Example 7: FIE-15G-ESA1 activated plants.

The FIE-15G-ESA1 transgene was transcriptionally activated by crossing FE #1 and FE #2 plants with the GAL4-VP16 two-component activation lines described in Table 3. Reciprocal crosses were carried out using FE #1 and FE #2 plants with each 2-component activation line. The seed produced in such a cross are referred to as F_1 seed. Thus, a first generation seed or plant produced by crossing FE #1 as the mother with J2592 as the pollen donor is referred to as F_1 (FE #1 x J2592). A second generation seed or plant produced by self pollination of F_1 (FE #1 x J2592) is referred to as F_2 (FE #1 x J2592). F_1 seed produced by crossing FE #1 and FE #2 with the activation lines described above were collected from mature siliques or seed pods and dried using standard *Arabidopsis* procedures. These siliques typically contained mature seed, abnormal seed and aborted ovules.

To analyze the effect of FIE-15G-ESA1 expression on *Arabidopsis* development F_1 seed and seed from control plants were germinated on agar plates containing 1x Murashige and Skoog (MS) salts and 1 percent sucrose using standard *Arabidopsis* procedures. Germinated seedlings were scored 8 days after plating for germination efficiency, the presence or absence of the activator gene (inferred from GFP reporter gene activity) and seedling phenotypes. After phenotyping, F_1 seedlings were transferred to soil at the four rosette leaf stage and then grown under standard *Arabidopsis* greenhouse conditions. Flowering plants were tested by PCR for the presence of the FIE-15G-ESA1 target gene and scored again for GFP expression.

When line J2592 was used as the activation line, 86 percent of the F_1 seeds germinated normally. F_1 seedlings and plants exhibited both vegetative and reproductive effects of FIE-15G-ESA1 activity. For example, cotyledons were observed to be incomplete, cupped, or missing in 30 percent of all seedlings analyzed. In some instances, extra cotyledons were observed. Hypocotyl development was perturbed in twelve percent of all F_1 seedlings analyzed. Finally, twenty-four percent of F_1 seedlings displayed stunted or missing petioles. Developmental abnormalities resulted in the loss of some seedlings from the study. These phenotypes were not observed in seedlings produced by selfing J2592, FE #1, or FE #2. Nor were these phenotypes observed in seedlings produced by crossing these parents with a wild type plant. The results indicate that activation of FIE-15G-ESA1 by J2592 is responsible for these diverse traits.

When activation line M0164 was used to activate FIE-15G-ESA1, ninety-seven percent of the F_1 (M0164 x FE #1) seed germinated successfully. Forty percent of F_1 seedlings analyzed showed vegetative defects including cotyledons that were incomplete, cupped, or missing. In some instances, extra cotyledons were observed. Thus, the F_1 seedling phenotypes induced using FIE-15G-ESA1 were not restricted to the J2592 activation line.

Reproductive phenotypes for F_1 plants containing activator and FIE-15G-ESA1 target genes were analyzed as described in Ohad, N., et al. (1999) *The Plant Cell* 11:407-415; and in Fischer, R.L., et al., (2001) US Patent 6,229,064. In brief,

developing siliques were sampled along the primary inflorescence proximal to distal relative to the rosette leaves. Within each silique, the seed were classified according to the color and the status of endosperm and embryo development. Since F_1 seed are the product of genetic crossing, each silique that is produced by an F_1 plant should contain a population of F_2 seed that segregate for the activator and target genes and any resulting phenotype. Thus, each silique contains a population of wild type seed that provide a developmental reference for staging seed development and phenotyping. Seed phenotypes were recorded at two stages of seed development: (i) when the majority of seed in a silique were at the mature seed stage of embryo development, and (ii) at the torpedo to walking stick stage of embryo development.

Effect of FIE-15G-ESA1 gene activity on seed development: F_2 seeds were produced by F_1 plants through self-pollination. F_2 (FE #1 x J2592) and F_2 (FE #2 x J2592) seed development was characterized using a Zeiss dissecting microscope and a Zeiss Axioskope microscope as described by Ohad, N., et al., (1999) *The Plant Cell* 11:407-415 using standard *Arabidopsis* procedures.

Activation of FIE-15G-ESA1 by J2592 altered embryo and seed development as shown in Table 4. Self-pollinated F_1 (FE #1 x J2592) plants produced two classes of seed, (i) those exhibiting normal embryo and seed development, and (ii) those exhibiting abnormal seed and embryo development. Abnormal seed were found to contain an embryo whose development was arrested at the transition between heart and torpedo stages of development. By contrast, endosperm production was not arrested in abnormal seed but was greater than or equal to that observed in normal seed. Thus, FIE-15G-ESA1 was observed to alter the balance between endosperm and embryo development within the seed. Most abnormal seed abort and degenerate into shrunken seed. The percent abnormal to normal seed ranged from 25-62 percent (see Table 4). Similar results also were observed in F_1 (FE #2 x J2592) plants. Similar results were observed when the reciprocal cross (i.e., J2592 x FE #1) was performed. FIE-15G-ESA1 also was observed to alter seed development when J0661 was crossed with FE #1. By contrast, no abnormal seed were detected in F_1 plants produced by crossing Q2500, J2301 or J2921 with FE #1. In fact, more than 98 percent of seed from self-pollinated FE #1, FE #2 and J2592 parental lines had no

visually observable abnormalities. Thus, the effect of FIE-15G-ESA1 activity on seed development appears to be promoter dependent.

TABLE 4.

Segregation of seed phenotypes in developing siliques

F1(FE#1 X J2592) Plant #29

	Normal seeds	Shrunken aborted seeds
Total	225	683
Percent	25.6	74.4

F1(FE#1 X J2592) Plant #35

	Normal seeds	Shrunken aborted seeds
Total	264	760
Percent	25.8	74.2

F1(FE#1 X J2592) Plant #31

	Normal seeds	Shrunken aborted seeds
Total	264	760
	25.8	74.2

Segregation of seed phenotypes in mature siliques

F1(FE#1 X J2592) Plant #35

	Normal seeds	Shrunken aborted seeds
Total	361	136
Percent		37.7
STD*		3.8

F1(FE#1 X J2592) Plant #29

	Normal seeds	Shrunken aborted seeds
Total	355	115
Percent		32.40
STD		3.40

F1(FE#1 X J2592) Plant #31

	Normal seeds	Shrunken aborted seeds
Total	364	111
Percent		30.5
STD		2.1

F1(FE#1 X J2592) Plant #26

	Normal seeds	Shrunken aborted seeds
Total	379	151
Percent		39.8
STD		4.1

F1(FE#1 X J2592) Plant #37

	Normal seeds	Shrunken aborted seeds
Total	369	125
Percent		33.9
STD		2.7

F1(FE#1 X J2592) Plant #32

	Normal seeds	Shrunken aborted seeds
Total	308	192
Percent		62.3
STD		5.2

* STD = standard deviation

Example 8: MEA-15G-ESA1 activated plants.

The MEA-15G-ESA1 transgene was transcriptionally activated by crossing ME #1 with J2592, J0661 and Q2500 (see Table 5). Reciprocal crosses between ME #1 and each activation line also were made. F_1 seeds were collected at maturity and stored under standard conditions. To analyze the effect of MEA-15G-ESA1

expression on *Arabidopsis* development F_1 seed and seed from control plants were germinated on agar plates containing 1x MS salts and 1 percent sucrose. Subsequently, plants were phenotyped as described in Example 7. Mature plants were tested for the presence of MEA-15G-ESA1 by PCR.

When J2592 or Q2500 were crossed with ME #1 (pCRS304-5UAS-MEA-15G-ESA1 transformed plant #1) the F_1 seedlings exhibited vegetative phenotypes similar to those caused by FIE-15G-ESA1 in F_1 (FE #1 x J2592) and (FE #1 x M0164). For example, the cotyledons of F_1 seedlings were observed to be incomplete, cupped, or missing. Hypocotyl development also was perturbed. These phenotypes were not observed in seedlings produced by the self pollination of J2592, Q2500 or ME #1. Thus, activation of MEA-15G-ESA1 by J2592 and Q2500 is responsible for these vegetative developmental effects.

Table 5.

Segregation of seed phenotypes in developing siliques					
Cross ID	Line ID	# Green seed	# White seed	# Aborted ovules	Total seeds
F1(ME #1 x J2592)	Plant #1				
	AVG	17.4	6	14.4	37.8
	%	46.0	15.9	38.1	100
F1(ME #1 x J2592)	Plant #2				
	AVG	17.2	6.3	15.9	39.4
	%	43.7	16.0	40.4	100
F1(ME #1 x J2592)	Plant #3				
	AVG	23	0	15	38
	%	60.5	0.0	39.5	100
F1(ME #1 x J2592)	Plant #22				
	AVG	18.5	5.9	14.8	39.2
	%	47.2	15.1	37.8	100
F1(ME #1 x J2592)	Plant #24				
	AVG	24.6	0	15.5	40.1
	%	61.3	0.0	38.7	100
Control (GFP negative)	Plant #26				
	AVG	43.2	0	0.2	43.4

	%	99.5	0.0	0.5	100
Cross ID	Line ID	# Green seed	# White seed	# Aborted ovules	Total est.
F1(J2592 X ME #1)	Plant #14				
	AVG	24.2	0.1	14.8	39.1
	%	61.9	0.3	37.9	100
F1(J2592 X ME #1)	Plant #26				
	AVG	15	6.2	16	37.2
	%	40.3	16.7	43.0	100
F1(J2592 X ME #1)	Plant #34				
	AVG	15.9	6.7	15.2	37.8
	%	42.1	17.7	40.2	100
Cross ID	Line ID	# Green seed	# White seed	# Aborted ovules	Total est.
F1(ME #1 x J0661)	Plant #32				
	AVG	15.9	0	12.4	28.3
	%	56.2	0.0	43.8	
F1(J0661 x ME #1)	Plant #19				
	AVG	17.8	0	17.5	35.3
	%	50.4	0.0	49.6	
F1(J0661 x ME #1)	Plant #27				
	AVG	18.6	0	15.4	34
	%	54.7	0.0	45.3	
F1(J0661 x ME #1)	Plant #28				
	AVG	18.2	0.1	16.8	35.1
	%	51.9	0.3	47.9	

Example 9: Fertilization independent seed development.

To determine the frequency of post-fertilization seed abortion, siliques harvested at two weeks and at four weeks after self-pollination were dissected, and wild-type and aborted seeds were counted. To test for fertilization-independent development, flower buds from plants that had not yet begun to shed pollen (i.e., stage 12 plants) (see Smyth, D.R., et al., *Plant Cell*, 2: 755-761 (1990)) were opened, immature anthers were removed, and the flower bud was covered with a plastic bag. In some experiments, the silique was measured, dissected, and the number of seed-

like structures and degenerated ovules were counted after seven days. In some experiments, the silique was harvested and ovules and seed-like structures were phenotyped after 15 days.

When immature F_1(J2592 x FE #1) flowers were emasculated and allowed to develop, seed-like structures were observed that were filled with endosperm but contained no embryo. This occurred in roughly 40 percent of the siliques analyzed. Thus, activation of FIE-15G-ESA1 by J2592 also can induce fertilization independent endosperm and seed development.

Example 10: Profiling gene expression.

This example demonstrates the use of chimeric polypeptides for RNA expression profiling. Gene expression in developing flowers from F_1 (J2592 x FE #1) was compared to gene expression in flowers from activation line J2592 (see Table 3) and target line FE #1 using microarray expression analysis. All experiments were done in duplicate.

Sample preparation: Seeds of F_1 (J2592 x FE #1) plants were sterilized in 95% bleach for 1 minute and with 70% ethanol for 45 seconds and subsequently washed 5 times in sterile distilled deionized water and then plated on MS agar plates and left at 4 °C for 4 days to be vernalized. Plates were placed in growth chamber with 16 hr light/8 hr. dark, 23 °C, 14,500-15,900 LUX, and 70% relative humidity for germination and growth. Seedlings were PCR-genotyped for the presence of the transgene and analyzed using dissecting microscopy for GFP expression before they were transplanted individually into soil. Tissues harvested for RNA extraction consisted of compact terminal inflorescences. Each sample contained a population of sequentially produced and continuously developing flowers representing all stages of flower development from early floral primordial, to immature floral buds, to mature flowers up to and including two days after pollination. Samples were flash frozen in liquid nitrogen and stored at -80 °C until use. Total RNA was extracted using Qiagen RNeasy Kit with the protocol recommended by manufacture and the RNA was then dissolved in RNA-free water.

Approximately 10 µg of the each RNA sample was used for amplification using MessageAmp™ aRNA Kit provided by Ambion, Inc. Poly(A+) mRNA was isolated using standard procedures (Poly(A) Quick mRNA Isolation Kit (Stratagene, La Jolla, California), and 2 µg from each sample was used to generate labeled probes for hybridization to microarray slides containing *Arabidopsis* cDNA sequences. The *Arabidopsis* microarray contained nucleic acid features representing 10,000 different *Arabidopsis* genes. Hybridization experiments to detect differentially regulated genes were set up in pairs. For example, RNA from the F_1 (J2592 x FE #1) plant was compared to RNA from either the *Arabidopsis* activation line J2592 or the *Arabidopsis* transgenic line FE#1. Expression results are analyzed using standard software and procedures.

Slide preparation: Microarray technology provides the ability to monitor mRNA transcript levels of thousands of genes in a single experiment. These experiments simultaneously hybridize two differentially labeled fluorescent cDNA pools to glass slides that have been previously spotted with cDNA clones of the same species. Each arrayed cDNA spot will have a corresponding ratio of fluorescence that represents the level of disparity between the amount of respective mRNA species in the two sample pools. Thousands of polynucleotides can be spotted on one slide, and each experiment analyzes the expression pattern of thousands of mRNA species.

The microarray utilizes a chemically coated microscope slide, referred herein as a "chip" with numerous polynucleotide samples arrayed at a high density. The coating with chemicals such as Poly-L-lysine allows for spotting DNA at high density by providing a hydrophobic surface, reducing the spreading of spots of DNA solution arrayed on the slides. Glass microscope slides (Gold Seal #3010 manufactured by Gold Seal Products, Portsmouth, New Hampshire, USA) were coated with a 0.1%W/V solution of Poly-L-lysine (Sigma, St. Louis, Missouri) using the following protocol:

Slides were placed in slide racks (Shandon Lipshaw #121). The racks were then put in chambers (Shandon Lipshaw #121). Cleaning solution was prepared by dissolving 70g NaOH in 280 mL ddH2O. 420 mL 95% ethanol was added. The total volume was 700 mL (= 2 X 350 mL) and the solution was stirred until completely mixed. If the solution remained cloudy, ddH_2O was added until the solution cleared. The cleaning

solution was poured into chambers with slide racks, and the chambers were covered with glass lids. The solution was mixed on an orbital shaker for 2 hr. The racks were quickly transferred to fresh chambers filled with ddH_2O and were rinsed vigorously by plunging racks up and down. Rinses were repeated 4 times with fresh ddH_2O each time, to remove NaOH-ethanol. Poly-L-lysine solution was prepared by adding 70 mL poly-L-lysine stock solution to 70 mL tissue culture PBS in 560 mL double-distilled deionized water using plastic graduated cylinders and beakers. Slides were transferred to polylysine solution and shaken on an orbital shaker for 1 hr. The rack was transferred to a fresh chamber filled with ddH_2O, and was plunged up and down 5 times to rinse. The slides were centrifuged on microtiter plate carriers (paper towels were placed below the rack to absorb liquid) for 5 min. @ 500 rpm. The slide racks were transferred to empty chambers with covers, and were dried in a 45 °C oven for 10 min. The slides were stored in a closed plastic slide box in the dark. Normally, the surface of lysine coated slides was not very hydrophobic immediately after this process, but became increasingly hydrophobic with storage. A hydrophobic surface helped ensure that spots did not run together while printing at high densities. After they aged for 10 days to a month the slides were ready for use. Stored slides that developed opaque patches, visible when held to the light, can result in high background hybridization from the fluorescent probe and were not used.

PCR amplification of cDNA clones: Polynucleotides were amplified from *Arabidopsis* cDNA clones using one insert specific primer and one common primer that hybridized to the cloning site. The resulting 100 µl PCR reactions were purified with Qiaquick 96 PCR purification columns (Qiagen, Valencia, California, USA) and eluted in 30 µL of 5mM Tris. 8.5 µL of the elution were mixed with 1.5 µL of 20X SSC to give a final spotting solution of DNA in 3X SSC. The concentrations of DNA generated from each clone varied between 10-100 ng/µl, but were usually about 50 ng/µl.

Arraying PCR products on slides: Purified PCR products were spotted onto poly-L-Lysine coated glass slides using an arrangement of quill-tip pins (ChipMaker 3 spotting pins; Telechem International, Inc., Sunnyvale, California, USA) and a robotic arrayer (PixSys 3500, Cartesian Technologies, Irvine, California, USA). Approximately 0.5 nl of a prepared PCR product was spotted at each location to produce spots having a diameter of about 100 µm. Spot were spaced 180 µm to 210 µm center-to-center. Printing was

conducted in a chamber with relative humidity set at 50%. Slides containing maize sequences were purchased from Agilent Technology (Palo Alto, CA 94304).

Slide processing: After arraying, slides were processed through a series of steps prior to hybridization: rehydration, UV cross-linking, blocking and denaturation. Slides were rehydrated by placing them over a beaker of warm (55 °C) water (DNA face down), for 2-3 sec to distribute the DNA evenly within the spots, and then snap dried on a hot plate (DNA side face up). The DNA was cross-linked to the slides by UV irradiation (60-65mJ; 2400 Stratalinker, Stratagene, La Jolla, California, USA). A blocking step was performed to modify remaining free lysine groups, and hence minimize their ability to bind labeled probe DNA. To achieve this, the arrays were placed in a slide rack. An empty slide chamber was left ready on an orbital shaker. The rack was bent slightly inwards in the middle, to ensure the slides would not run into each other while shaking. The blocking solution was prepared as follows:

Three 350-ml glass chambers (with metal tops) were set to one side, and a large round Pyrex dish with dH_2O was placed ready in the microwave. At this time, 15ml sodium borate was prepared in a 50 ml conical tube. 6 g succinic anhydride was dissolved in about 325-350 mL 1-methyl-2-pyrrolidinone. Rapid addition of reagent was important. Immediately after the last flake of the succinic anhydride dissolved, 15-mL sodium borate was added. Immediately after the sodium borate solution mixed in, the solution was poured into an empty slide chamber. The slide rack was plunged rapidly and evenly in the solution and was vigorously shaken up and down for a few seconds, making sure slides never left the solution. It was mixed on an orbital shaker for 15-20 min. Meanwhile, the water in the Pyrex dish (enough to cover slide rack) was heated to boiling. Following this, the slide rack was gently plunged into 95 °C water for 2 min. The slide rack then was plunged 5times in 95% ethanol. The slides and rack were centrifuged for 5 min. at 500 rpm. Slides were loaded quickly and evenly onto the carriers to avoid streaking, and were used immediately or were stored in a slide box.

Hybridization: The hybridization process began with the isolation of mRNA from the two tissues followed by their conversion to single stranded cDNA (see *"Generation of probes for hybridization"*, below). The cDNA from each tissue was independently labeled with a different fluorescent dye and then both samples were pooled together. This

final differentially labeled cDNA pool was then placed on a processed microarray and allowed to hybridize (see *"Hybridization and wash conditions"*, below).

<u>*Preparation of Yeast control mRNA:*</u> Plasmid DNA was isolated from the following yeast clones using Qiagen filtered maxiprep kits (Qiagen, Valencia, California): YAL022c(Fun26), YAL031c(Fun21), YBR032w, YDL131w, YDL182w, YDL194w, YDL196w, YDR050c and YDR116c. Plasmid DNA was linearized with either *Bsr*BI (YAL022c(Fun26), YAL031c(Fun21), YDL131w, YDL182w, YDL194w, YDL196w, YDR050c) or *Afl*III (YBR032w, YDR116c).

The following solution was incubated at 37 °C for 2 hours: 17 µl isolated yeast insert DNA (1 µg), 20 µl 5X buffer, 10 µl 100 mM DTT, 2.5 µl (100 U) RNasin, 20 µl 2.5 mM (ea.) rNTPs, 2.7 µl (40U) SP6 polymerase and 27.8 µl RNase-free deionized water. Two µl (2 U) Ampli DNase I was added and the incubation continued for another 15 min. Ten µl 5M NH_4OAC and 100 µl phenol:chloroform:isoamyl alcohol (25:24:1) were added, and the solution was vortexed and centrifuged to separate the phases. To precipitate the RNA, 250 µl ethanol was added and the solution was incubated at -20 °C for at least one hour. The sample was then centrifuged for 20 min. at 4 °C at 14,000-18,000 X g, the pellet was washed with 500 µl of 70% ethanol, air dried at room temperature for 10 min. and resuspended in 100 µl of RNase-free deionized water. The precipitation procedure was repeated one time.

Alternatively, after the two-hour incubation, the solution was extracted with phenol/chloroform once before adding 0.1 volume 3M sodium acetate and 2.5 volumes of 100% ethanol. The solution was centrifuged at 15,000rpm, 4 °C for 20 min. and the pellet resuspended in RNase-free deionized water. The DNase I treatment was carried out at 37°C for 30 min. using 2 U of Ampli DNase I in the following reaction condition: 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$. The DNase I reaction was then stopped with the addition of NH_4OAC and phenol:chloroform:isoamyl alcohol (25:24:1), and RNA isolated as described above.

0.15-2.5 ng of the *in vitro* transcript RNA from each yeast clone was added to each plant mRNA sample prior to labeling to serve as positive (internal) probe controls.

Generation of labeled probes for hybridization from first-strand cDNA:

Hybridization probes were generated from isolated mRNA using an Atlas™ Glass Fluorescent Labeling Kit (Clontech Laboratories, Inc., Palo Alto, California, USA). This entails a two step labeling procedure that first incorporates primary aliphatic amino groups during cDNA synthesis and then couples fluorescent dye to the cDNA by reaction with the amino functional groups. Briefly, 5 μg of oligo(dT)$_{18}$ primer d(TTTTTTTTTTTTTTTTTTV) was mixed with Poly A+ mRNA (1.5 - 2 μg mRNA isolated using the Qiagen Oligotex mRNA Spin-Column protocol or-the Stratagene Poly(A) Quik mRNA Isolation protocol (Stratagene, La Jolla, California, USA) in a total volume of 25 μl. The sample was incubated in a thermocycler at 70°C for 5 min., cooled to 48 °C and 10 μl of 5X cDNA Synthesis Buffer (kit supplied), 5 μl 10X dNTP mix (dATP, dCTP, dGTP, dTTP and aminoallyl-dUTP; kit supplied), 7.5 μl deionized water and 2.5 μl MMLV Reverse Transcriptase (500 U) added. The reaction was then incubated at 48 °C for 30 min., followed by a 1 hr incubation at 42 °C. At the end of the incubation, the reaction was heated to 70 °C for 10 min., cooled to 37 °C and 0.5 μl (5 U) RNase H added, before incubating for 15 min. at 37 °C. The solution was vortexed for 1 min. after the addition of 0.5 μl 0.5 M EDTA and 5 μl of QuickClean Resin (kit supplied) then centrifuged at 14,000-18,000 X g for 1 min. After removing the supernatant to a 0.45 μm spin filter (kit supplied), the sample was again centrifuged at 14,000-18,000 X g for 1 min., and 5.5 μl 3 M sodium acetate and 137.5 μl of 100% ethanol added to the sample before incubating at –20 °C for at least 1 hr. The sample was then centrifuged at 14,000-18,000 X g at 4 °C for 20 min., the resulting pellet washed with 500 μl 70% ethanol, air-dried at room temperature for 10 min. and resuspended in 10 μl of 2X fluorescent labeling buffer (kit provided). 10 μl each of the fluorescent dyes Cy3 and Cy5 (Amersham Pharmacia , Piscataway, New Jersey, USA); prepared according to Atlas™ kit directions of Clontech) were added and the sample incubated in the dark at room temperature for 30 min. to 1 hr.

The fluorescently labeled first strand cDNA was precipitated by adding 2 μl 3M sodium acetate and 50 μl 100% ethanol, incubated at -20°C for at least 2 hrs, centrifuged at 14,000-18,000 X g for 20 min., washed with 70% ethanol, air-dried for 10 min. and

dissolved in 100 µl of water. Alternatively, 3-4 µg mRNA, 2.5 (~8.9 ng of in vitro translated mRNA) µl yeast control and 3 µg oligo dTV (TTTTTTTTTTTTTTTTTT(A/C/G); Sequence ID No.: X) were mixed in a total volume of 24.7 µl. The sample was incubated in a thermocycler at 70°C for 10 min. before chilling on ice. To this, 8 µl of 5X first strand buffer (SuperScript II RNase H- Reverse Transcriptase kit from Invitrogen, Carlsbad, California 92008; cat no. 18064022), 0.8 µl of aa-dUTP/dNTP mix (50X; 25mM dATP, 25mM dGTP, 25mM dCTP, 15mM dTTP, 10mM aminoallyl-dUTP), 4 µl of 0.1 M DTT and 2.5 µl (500 U) of Superscript R.T.II enzyme (Stratagene) were added. The sample was incubated at 42 °C for 2 hours before a 10 °C mixture of 1M NaOH and 0.5 M EDTA was added. After a 15 minute incubation at 65 °C, 25 µl of 1 M Tris pH 7.4 was added. This was mixed with 450 µl of water in a Microcon 30 column before centrifugation at 11,000 X g for 12 min. The column was washed twice with 450 µl (centrifugation at 11,000 g for 12 min.) before eluting the sample by inverting the Microcon column and centrifuging at 11,000 X g for 20 seconds. Sample was dehydrated by centrifugation under vacuum and stored at -20 °C.

Each reaction pellet was dissolved in 9 µl of 0.1 M carbonate buffer (0.1M sodium carbonate and sodium bicarbonate, pH=8.5-9) and 4.5 µl of this was placed in two microfuge tubes. 4.5 µl of each dye (in DMSO) was added, and the mixture was incubated in the dark for 1 hour. 4.5 µl of 4 M hydroxylamine was added and the mixture was again incubated in the dark for 15 min.

Irrespective of the method used for probe generation, the probe was purified using a Qiagen PCR cleanup kit (Qiagen, Valencia, California, USA), and eluted with 100 ul EB (kit provided). The sample was loaded on a Microcon YM-30 (Millipore, Bedford, Massachusetts, USA) spin column and concentrated to 4-5 ul in volume. Probes for the maize microarrays were generated using the Fluorescent Linear Amplification Kit (cat. No. G2556A) from Agilent Technologies (Palo Alto, CA).

Hybridization Conditions: Labeled probe was heated at 95 °C for 3 min. and was then chilled on ice. Then, 25 µl of the hybridization buffer which was warmed at 42 °C was added to the probe and was mixed by pipetting to give a final concentration of: 50% formamide, 4x SSC, 0.03% SDS, 5x Denhardt's solution, and 0.1 µg/ml single-stranded

salmon sperm DNA. The probe was kept at 42 °C. Prior to hybridization, the probe was heated for 1 min., added to the array, and then covered with a glass cover slip. Slides were placed in hybridization chambers (Telechem International, Sunnyvale, California) and incubated at 42 °C overnight.

Washing conditions: Slides first were washed in 1x SSC + 0.03% SDS solution at room temperature for 5 min. Slides then were washed in 0.2x SSC at room temperature for 5 min. Slides finally were washed in 0.05x SSC at room temperature for 5 min. Slides then were spun at 800 x g for 2 min. to dry. They were then scanned.

Scanning of slides: Chips were scanned using a ScanArray 3000 or 5000 (General Scanning, Watertown, Massachusetts, USA). The chips were scanned at 543 nm and at 633 nm at a resolution of 10 µm to measure the intensity of the two fluorescent dyes incorporated into the samples hybridized to the chips.

Data extraction and analysis: The images generated by scanning slides consisted of two 16-bit TIFF images representing the fluorescent emissions of the two samples at each arrayed spot. These images were quantified and processed for expression analysis using Imagene™ (Biodiscovery, Los Angeles, California, USA) data extraction software. Imagene™ output was using the Genespring ™ (Silicon Genetics, San Carlos, California, USA) analysis software. In Genespring ™, the data was imported using median pixel intensity measurements derived from Imagene™ output. Ratio calculation and normalization were conducted using Genespring™. Normalization was achieved by parsing the data into 32 groups, each of which represented one of the 32 pin printing regions on the microarray. Each group consisted of about 360 to 550 spots, and was independently normalized by setting the median of ratios to one and multiplying ratios by the appropriate factor.

Results: Among the ten thousand genes represented on the DNA chip, the expression ratio of 152 genes (1.52%) was found to be reduced at least 2-fold in F_1 (J2592 x FE #1) floral tissue when compared to floral tissue from FE #1. Similarly, the expression ratios of 63 genes (0.63%) were found to be down at least 2-fold in F_1 (J2592 x FE #1) floral tissue when compared to floral tissue from J2592. By contrast, the expression ratio of 227 genes (2.27%) was increased more than 2-fold in F_1 (J2592 x FE

#1) floral tissue when compared to floral tissue from FE #1. Similarly, 39 genes (0.39%) were found to be up at least 2-fold in F_1 (J2592 x FE #1) compared to J2592 floral tissue.

Example 11: Analysis of FIE-15G-ESA1 Activated Plants

The FIE-15G-ESA1 transgene is transcriptionally activated by crossing female FE plants containing a FIE-15G-ESA1 transgene to enhancer trap HAP1-VP16 lines that display cell and tissue specific GFP accumulation in vegetative and reproductive organs. FE plants are crossed with four different activation lines. A different enhancer is present in each of the lines and confers expression of the HAP1-VP16 transcription activator, as well as the GFP, in a different set of tissues. The amino acid sequence of the HAP1 portion of the HAP1-VP16 transcription activator is that of the yeast HAP1 gene. The activity of each enhancer-trap line is inferred from the GFP fluorescence.

At maturity, F_1 seeds are collected and stored under standard conditions. A reciprocal cross is also made, in which FE plants are used as males.

F_1 seeds are germinated and allowed to self-pollinate. After pollination, some of the embryos and seeds developing on F_1 plants are examined under a microscope. Mature seed also are analyzed as described in Example 7. Seedlings are scored for GFP expression and tested for the presence of FIE-15G-ESA1 by PCR. Phenotypic traits are analyzed as described in Example 7.

REFERENCE LIST

Aalfs, J.D. and Kingston, R.E. (2000) *Trends Biochem Sci.* 25:548-55.

Cress W.D. and Seto, E. (2000) *J Cell Physiol.* 184:1-16.

Amedeo, P., Habu, Y., Afsar, K., Scheid, O.M. and Paszkowski, J. (2000) *Nature* 405:203-206.

Cao, X., Springer, N.M., Muszynski, M.G., Phillips, R.L., Kaeppler, S. and Jacobsen, S.E. (2000) *Proc. Natl. Acad. Sci. USA* 97:4979-4984.

Butler, K.H., Danilevskaya, O., Miao, G-H., Morgante, M., Sakai, H., Simmons, C.R., Weng, Z. (2001) PCT WO0116325.

Davie, J.R. and Moniwa, M., (2000) *Crit Rev Eukaryot Gene Expr.* 10:303-325.

Dellaporta, S.L. and Chen, J. (2000) US Patent 6,011,200.

Farkas, G. et al. (2000) *Gene.* 253:117-36.

Fischer, R.L., Ohad, N., Kiyosue, T., Yadegari, R., Margossian, L., Harada, J.J., Goldberg, R.B. (2001) US Patent 6,229,064.

Flaus, A. and Owen-Hughes, T. (2001) *Curr Opin Genet Dev.* 11:148-54.

Fry, C.J. and Peterson, C.L. (2001) *Curr Biol.* 11:R185-97.

Gregory, P.D. et al. (2001) *Exp Cell Res.* 265:195-202.

Grossniklaus, U., Vielle-Calzada, J-P., Hoeppner, M.A. and Gagliano, W.B. (1998) *Science* 280: 446-450.

Grossniklaus, U. and Vielle-Calzada, J-P. (2001) US Patent 6,239,327

Haseloff, J., Siemering, K.R., Prasher, D. C. and Hodge, S. (1997) *Proc. Natl. Acad. Sci. USA* 94:2122-2127

Jackson, J.P., Lindroth, A.M., Cao, X. and Jacobsen, S.E. (2002) *Nature* 416:556-560.

Jeddeloh, J.A., Stokes, T.L. and Richards, E.J. (1999) *Nature Genetics* 22:94-97.

Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., Katz, A., Margossian, L., Harada, J.J., Goldberg, R.B. and Fischer, R.L. (1999) *Proc. Natl. Acad. Sci. USA* 96: 4186-4191.

Krebs, J.E. and Peterson, C.L. (2000) *Crit Rev Eukaryot Gene Expr.* 10:1-12.

Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, Jacobsen SE. (2001) *Science* 292:2077-2080.

Liu, Z-B. and Odell, J.T. (1999) US Patent # 5,968,793.

Luo, M., Bilodeau, P., Dennis, E. S., Peacock, W. J. and Chaudhury, A. (2000) *Proc. Natl. Acad. Sci. USA* 97:10637-10642.

Muller, C. and Leutz, A. (2001) *Curr Opin Genet Dev.* 11:167-74.

Ohad, N., Yadegari, R., Margossian, L., Hannon, M., Michaeli, D., Harada, J.J., Goldberg, R.B. and Fischer, R.L. (1999) *The Plant Cell* 11:407-415.

Richards, E.J., Jeddeloh, J.A. US patent 6,153,741.

Roseman, R.R., Morgan, K., Mallin, D.R., Roberson, R., Parnell, T.J., Bornemann, D.J., Simon, J.A. and Geyer, P.K. (2001) *Genetics* 158:291-307.

Springer, N.M., Danilevskaya, O.N., Hermon, P., Helentjaris, T.G., Phillips, R.L., Kaeppler, H.F., and Kaeppler, S.M. (2002) *Plant Physiology* 128:1332-1345.

Vielle-Calzada, J-P., Thomas, J., Spillane, C., Coluccio, A., Hoeppner, M. A. and Grossniklaus, U. (1999) *Genes & Dev.* 13:2971-2982.

Vinkenoog, R., Spielman, M., Adams, S. Fischer, R. L., Dickinson, H.G. and Scott, R.J. (2000) *The Plant Cell* 12:2271-2282.

Wu, K., Tian, L., Malik, K., Brown, D. and Miki, B. (2000) *The Plant Journal* 22:19-27.

Wu, K., Miki, B.L.A., Tian, L., Brown, D.C.W. (2001) European Patent Application 1094112A2.

Table 6: HAT Polypeptide Sequences

Arabidopsis ESA1-like

MGSSANTETNGNAPPPSSNQKPPATNGVDGSHPPPPPLTPDQAIIESDPSKKRKMGMLPL
EVGTRVMCRWRDGKHHPVKVIERRRIHNGGQNDYEYYVHYTEFNRRLDEWTQLDQLDLDS
VECAVDEKLEDKVTSLKMTRHQKRKIDETHIEGHEELDAASLREHEEFTKVKNISTIELG
KYEIETWYFSPFPPEYNDCVKLFFCEFCLNFMKRKEQLQRHMXKCDLKHPPGDEIYRSGT
LSMFEVDGKKNKVYAQNLCYLAKLFLDHKTLYYDVDLFLFYVLCECDDRGCHMVGYFSKE
KHSEEAYNLACILTLPSYQRKGYGKFLIAFSYELSKKEGKVGTPXKTLVGSRLTKLQRLL
DSCSIRNLEKT

Maize HAC000003

MDSHSSHLNAANRSRSSQTPSPSHSASASVTSSLHKRKLAATTAANAAASEDHAPPSSSF
PPSSFSADTRDGALTSNDELESISARGADTDSDPDESEDIVVDDDEDEFAPEQDQDSSIR
TFTAARLDSSSGVNGSSRNTKLKTESSTVKLESSDGGKDGGSSVVGTGVSGTVGGSSISG
LVPKDESVKVLAENFQTSGAYIAREEALKREEQAGRLKFVCYSNDSIDEHMMCLIGLKNI
FARQLPNMPKEYIVRLLMDRKHKSVMVLRGNLVVGGITYRPYHSQKFGEIAFCAITADEQ
VKGYGTRLMNHLKQHARDVDGLTHFLTYADNNAVGYFVKQEIPQSFTSKSSVSTLSYQGF
TKEIYLEKDVWHGFIKDYDGGLLMECKIDPKLPYTDLSSMIRQQRKAIDERIRELSNCQN
VYPKIEFLKNEAGIPRKIIKVEEIRGLREAGWTPDQWGHTRFKLFNGSADMVTNQKQLNA
LMRALLKTMQDHADAWPFKEPVDSRDVPDYYDIIKDPIDLKVIAKRVESEQYYVTLDMFV
ADARRMFNNCRTYNSPDTIYYKCATRLETHFHSKVQAGLQSGAKSQ

Arabidopsis HAT1

MSVHVKEEPVLVPNCDVENTELAVFNGNGESELENFGTCVDEITDRVNQLEQKVVEVEHF
YSTKDGAAQTNTSKSNSGGKKIAISQPNNSKGNSAGKEKSKGKHVSSPDLMRQFATMFRQ
IAQHKWAWPFLEPVDVKGLGLHDYYKVIEKPMDLGTIKKKMESSEYSNVREIYADVRLVF
KNAMRYNEEKEDVYVMAESLLEKFEEKWLLIMPKLVEEEKKQVDEEAEKHANKQLTMEAA
QAEMARDLSNELYEIDLQLEKLRESVVQRCRKLSTQEKKGLSAALGRLSPEDLSKALKMV
SESNPSFPAGAPEVELDIDVQTDVTLWRLKVFVQEALKAANKSSGGTNAQNNNNTGTGEI
NKNNAKRRREISDAINKASIKRAKKA

Table 7: CAP/HDAC Gene and Polypeptide Sequences

MEA, GI:3089625

MEKENHEDDG EGLPPELNQI KEQIEKERFL HIKRKFELRY IPSVATHASH HQSFDLNQPA AEDDNGGDNK SLLSRMQNPL
RHFSASSDYN SYEDQGYVLD EDQDYALEED VPLFLDEDVP LLPSVKLPIV EKLPRSITWV FTKSSQLMAE SDSVIGKRQI
YYLNGEALEL SSEEDEEDEE EDEEEIKKEK CEFSEDVDRF IWTVGQDYGL DDLVVRRALA KYLEVDVSDI LERYNELKLK
NDGTAGEASD LTSKTITTAF QDFADRRHCR RCMIFDCHMH EKYEPESRSS EDKSSLFEDE DRQPCSEHCY LKVRSVTEAD
HVMDNDNSIS NKIVVSDPNN TMWTPVEKDL YLKGIEIFGR NSCDVALNIL RGLKTCLEIY NYMREQDQCT MSLDLNKTTQ
RHNQVTKKVS RKSSRSVRKK SRLRKYARYP PALKKTTSGE AKFYKHYTPC TCKSKCGQQC PCLTHENCCE KYCGCSKDCN
NRFGGCNCAI GQCTNRQCPC FAANRECDPD LCRSCPLSCG DGTLGETPVQ IQCKNMQFLL QTNKKILIGK SDVHGWGAFT
WDSLKKNEYL GEYTGELITH DEANERGRIE DRIGSSYLFT LNDQLEIDAR RKGNEFKFLN HSARPNCYAK LMIVRGDQRI
GLFAERAIEE GEELFFDYCY GPEHADWSRG REPRKTGASK RSKEARPAR

FIS2, GI:4185501

MTLKAEVVEN FSCPFCLIPC GGHEGLQLHL KSSHDAFKFE FYRAEKDHGP EVDVSVKSDT IKFGVLKDDV GNPQLSPLTF
CSKNRNQRRQ RDDSNNVKKL NVLLMELDLD DLPRGTENDS THVNDDNVSS PPRAHSSEKI SDILTTTQLA IAESSEPKVP
HVNDGNVSSP PRAHSSAEKN ESTHVNDDDD VSSPPRAHSL EKNESTHVNE DNISSPPKAH SSKKNESTHM NDEDVSFPPR
TRSSKETSDI LTTTQPAIVE PSEPKVRRGS RRKQLYAKRY KARETQPAIA ESSEPKVLHV NDENVSSPPE AHSLEKASDI
LTTTQPAIAE SSEPKVPHVN DENVSSTPRA HSSKKNKSTR KNVDNVPSPP KTRSSKKTSD ILTTTQPTIA ESSEPKVRHV
NDDNVSSTPR AHSSKKNKST RKNDDNIPSP PKTRSSKKTS NILTRTQPAI AESEPKVPHV NDDKVSSTPR AHSSKKNKST
HKKDDNASLP PKTRSSKKTS DILATTQPAK AEPSEPKVTR VSRRKELHAE RCEAKRLERL KGRQFYHSQT MQPMTFEQVM
SNEDSENETD DYALDISERL RLERLVGVSK EEKRYMYLWN IFVRKQRVIA DGHVPWACEE FAKLHKEEMK NSSSFDWWWR
MFRIKLWNNG LICAKTFHKC TTILLSNSDE AGQFTSGSAA NANNQQSMEV DE

FIE, GI:4567095

MSKITLGNES IVGSLTPSNK KSYKVTNRIQ EGKKPLYAVV FNFLDARFFD VFVTAGGNRI TLYNCLGDGA ISALQSYADE
DKEESFYTVS WACGVNGNPY VAAGGVKGII RVIDVNSETI HKSLVGHGDS VNEIRTQPLK PQLVITASKD ESVRLWNVET
GICILIFAGA GGHRYEVLSV DFHPSDIYRF ASCGMDTTIK IWSMKEFWTY VEKSFTWTDD PSKFPTKFVQ FPVFTASIHT
NYVDCNRWFG DFILSKSVDN EILLWEPQLK ENSPGEGASD VLLRYPVPMC DIWFIKFSCD LHLSSVAIGN QEGKVYVWDL
KSCPPVLITK LSHNQSKSVI RQTAMSVDGS TILACCEDGT IWRWDVITK

Multi sex combs (mxc), AT5g46250

MESPSISDAVPLHAPEDATADFSQPQSPLHEVDSFPVTESSDDVVVNVSEIPNLSPSDDDFDHERNSGEDRDQDHGENPVETDGVVVPID
ELNQKIIRQVEYYFSDENLPTDKFLLNAMKRNKKGFVPISTIATFHKMKKLTRDHALIVSALKESSFLVVSADEKKVKRLSPLPEIRDPK
IFTVLVENLPEDHSNENIREIFGKAGSIKSVSICDPNAVEESEKGGKKENFIRTRLHAFVEYETVEAAEKAAATLNNEQDWRNGLRVKLL
EQAAGKFAQRRPARREVDKEKDTTGRVHDQTGGEKNKKTREHQNHRLHHSDNPADDDGGNHQKDKNGNKGRVVGQGRRQNHQGGNGIGHG
TASSSSHPNYHPVEVSKRPPGPRMPDGTRGFTMGRGKAIPPPTSTQTSHEV

Arabidposis TS01-like, GI:7767427

MDTPEKSETQ IGTPVSKLKV EDSPVFSYIC NLSPIKTIKP IPITCPLSSL NYASPPSVFT SPHAVSHKES RFRSQKDVSA
SKEVGEEEAL VGSEPEQSYK NDCNTPRVLN DVKDNGCGKD LQVMMDNVKK KSDTPDWETL IAATTELIYG SPRESEAFSC
LLKKTSNSEA RLRGSITATS VAVTNTDVVN NESESVDALS ILHRGVRRRC LDFEVKGNNQ QTLGESSSSC VVPSIGLHLN
TIAMSSKDKN VANEYSFSGN IKVGVQSSLT PVLHSQHDIV RENESGKDSG QIIEVVPKSL ASVDLTPISP KKKRRKSEQS
GEGDSSCKRC NCKKSKCLKL YCECFAAGFY CIEPCSCINC FNKPIHKDVV LATRKQIESR NPLAFAPKVI RNSDSIIEVG
EDASKTPASA RHKRGCNCKK SNCLKKYCEC YQGGVGCSIN CRCEGCKNAF GRKDGSLFEQ DEENETSGTP GTKKTQQNVE
LFKPAAPPST PIPFRQPLAQ LPISSNNRLL PPQSHFHHGA IGSSSSGIYN IRKPDMSLLS HSRIETITED IDDMSENLIH
SPITTLSPNS KRVSLSHLDS PESTPWRRNG EGRNLIRSFP TFPSLTPHH

Sin3, F3I6.12

MVGGGSAQKLTTNDALAYLKAVKDKFQDQRGKYDEFLEVMKNFKSQRVDTAGVITRVKELFKGHQELILGFNTFLPKGFEITLQPEDGQP
PLKKRVEFEEAISFVNKIKTRFQGDDRVYKSFLDILNMYRRDSKSITEVYQEVAILFRDHSDLLVEFTHFLPDTSATASIPSVKTSVRER
GVSLADKKDRIITPHPDHDYGTEHIDQDRERPIKKENKEHMRGTNKENEHRDARDFEPHSKKEQFLNKKQKLHIRGDDPAEISNQSKLSG
AVPSSSTYDEKGAMKSYSQDLAIVDRVKEKLNASEYQEFLRCLNLFSKEIISRPELQSLVGNLIGVYPDLMDSFIEFLVQCEKNEKRQIC
NLLNLLAEGLLSGILTKKSLWSEGKYPQPSLDNDRDQEHKRDDGLRDRDHEKERLEKAAANLKWAKPISELDLSNCEQCTPSYRLLPKNY
PISIASQKTEIGKLVLNDHWVSVTSGSEDYSFSHMRKNQYEESLFKCEDDRFELDMLLESVNSTTKHVEELLTKINSNELKTNSPIRVED
HLTALNLRCIERLYGDHGLDVMDVLKKNVSLALPVILTRLKQKQEEWARCRSDFDKVWAEIYAKNYYKSLDHRSFYFKQQDSKKLSMKAL
LAEIKEITEKKREDDSLLAFAAGNRLSISPDLEFDYPDHDLHEDLYQLIKYSCAEMCSTEQLDKVMKIWTTFVEQIFGVPSRPQGAEDQE
DVVKSMNQNVKSGSSSAGESEGSPHNYASVADSRRSKSSRKANEHSQLGQTSNSERDGAAGRTSDALCETAQHEKMLKNVVTSDEKPESK
QAVSIERAHDSTALAVDGLLDQSNGGSSIVHMTGHCNNNLKPVTCGTELELKMNDGNGPKLEVGNKKLLTNGIAVEITSDQEMAGTSKVE

REEGELSPNGDFEEDNFAVYAKTDFETFSKANDSTGNNISGDRSREGEPSCLETRAENDAEGDENAARSSEDSRNEYENGDVSGTESGGG
EDPEDDLDNNNKGESEGEAECMADAHDAEENGSALPVSARFLLHVKPLVKYVPSAIALHDKDKDSLKNSQVFYGNDSFYVLFRLHRILYE
RILSAKVNSSSPEGKWRTSNTKNPTDSYARFMTALYNLLDGTSDNAKFEDDCRAIIGTQSYILFTLDKLIHKFIKHLQVVVADEMDNKLL
QLYFYEKSRRPETIFDAVYYDNTRVLLPDENIYRIECRLSTPAKLSIQLMCNGLDKPDVTSVSIDPTFAAYLHNDFLSIQPNAREDRRIY
LNR

Sin3, GI:2829870
MVGGGSAQKL TTNDALAYLK AVKDKFQDQR GKYDEFLEVM KNFKSQRVDT AGVITRVKEL FKGHQELILG FNTFLPKGFE
ITLQPEDGQP PLKKRVEFEE AISFVNKIKT RFQGDDRVYK SFLDILNMYR RDSKSITEVY QEVAILFRDH SDLLVEFTHF
LPDTSATASI PSVKTSVRER GVSLADKKDR IITPHPDHDY GTEHIDQDRE RPIKKENKEH MRGTNKENEH RDARDFEPHS
KKEQFLNKKQ KLHIRGDDPA EISNQSKLSG AVPSSSTYDE KGAMKSYSQD LAIVDRVKEK LNASEYQEFL RCLNLFSKEI
ISRPELQSLV GNLIGVYPDL MDSFIEFLVQ CEKNEKRQIC NLLNLLAEGL LSGILTKKSL WSEGKYPQPS LDNDRDQEHK
RDDGLRDRDH EKERLEKAAA NLKWAKPISE LDLSNCEQCT PSYRLLPKNY PISIASQKTE IGKLVLNDHW VSVTSGSEDY
SFSHMRKNQY EESLFKCEDD RFELDMLLES VNSTTKHVEE LLTKINSNEL KTNSPIRVED HLTALNLRCI ERLYGDHGLD
VMDVLKKNVS LALPVILTRL KQKQEEWARC RSDFDKVWAE IYAKNYYKSL DHRSFYFKQQ DSKKLSMKAL LAEIKEITEK
KREDDSLLAF AAGNRLSISP DLEFDYPDHD LHEDLYQLIK YSCAEMCSTE QLDKVMKIWT TFVEQIFGVP SRPQGAEDQE
DVVKSMNQNV KSGSSSAGES EGSPHNYASV ADSRRSKSSR KANEHSQLGQ TSNSERDGAA GRTSDALCET AQHEKMLKNV
VTSDEKPESK QAVSIERAHD STALAVDGLL DQSNGGSSIV HMTGHCNNNL KPVTCGTELE LKMNDGNGPK LEVGNKKLLT
NGIAVEITSD QEMAGTSKVE REEGELSPNG DFEEDNFAVY AKTDFETFSK ANDSTGNNIS GDRSREGEPS CLETRAENDA
EGDENAARSS EDSRNEYENG DVSGTESGGG EDPEDDLDNN NKGESEGEAE CMADAHDAEE NGSALPVSAR FLLHVKPLVK
YVPSAIALHD KDKDSLKNSQ VFYGNDSFYV LFRLHRILYE RILSAKVNSS SPEGKWRTSN TKNPTDSYAR FMTALYNLLD
GTSDNAKFED DCRAIIGTQS YILFTLDKLI HKFIKHLQVV VADEMDNKLL QLYFYEKSRR PETIFDAVYY DNTRVLLPDE
NIYRIECRLS TPAKLSIQLM CNGLDKPDVT SVSIDPTFAA YLHNDFLSIQ PNAREDRRIY LNR

Arabidopsis MeCP2, GI:2827551
MNLKKSRSEN SSVASSGSKI EEQTEKSAEP TTIKVQKKAG TPGRSIDVFA VQCEKCMKWR KIDTQDEYED IRSRVQEDPF
FCKTKEGVSC EDVGDLNYDS SRTWVIDKPG LPRTPRGFKR SLILRKDYSK MDAYYITPTG KKLKSRNEIA AFIDANQDYK
YALLGDFNFT VPKVMEETVP SGILSDRTPK PSRKFLSGKM QGGGGRDPFG GGFGGPFGGF GGGSFGGFGR GSFGGFGGPN
GPPSLMSNFF GGRDPFDDPF FTQPFGGGMF QSNFFGPSMN PFAEMHRLPQ GFIENNQPPG PSRSRGPVIE EIDSDDEKEG
EGDKEKKGSL GKHGRSSSEA ETEDARVRER RNRQMQSMNV NAERRNREMQ NMNVNAERRN PQMQNMNVNA MVNNGQWQPQ
TGSYSFQSST VTYGGQNGNY YTSSKTRRTG SDGGHTVARK LNSDGRVDTT QTLHNLNEGG LVNREQPMLL PSTDPSPSHA
RAESSRRPKA AMNLIPILAI AVASAAFLSE LVSMSLPESI WRMMTPKAKI SVFSVNFPVV TYSGAKYPIA ALVLSAEKTI
SSARRL

Corn MBD1, GI:13936238
MTTGSTPGSAPSQRKRNSTKDSVALYAVQCYKCYKWSTVPKEEFETLRENFTKDPWFCSRRPDSSCEDDADIEYDSSRIWVLDKPNIPKP
PPETERLVVMRGDYSKMDTYYVMPNGKRARCAGDVDKFLEANPEYKDRISASDFSFAPPKVVEETVSHNPAWQAAKAKKQEKAEAQK

Corn MBD1, GI:13936310
MPAPDGWTKKFTPQRGGRSEIVFVSPTGEEIKNKRQLSQYLKAHPGGPAASDFDWGTGDTPRRSARISEKVKVFDSPEGEKIPKRSRNSS
GRKGRQGKKEAPETEEAKDAETGQDAPSEDGTKETDVEMKPAEEAKEAPTETDDAEKAADKADDTPAPAPMEEDEKETEKPAEAVVAPLA
QSEEKKEDAKPDEPEAVAPAPVSNPTENSAPAPAEPAAVPAPVPETESVAEPAAVLAPAPETKPDAKPAAVPAPAPENKPDAEPAAAAAP
VPDTKSVAEPAAAPAPDTKSVAEPAAAAPVPETKLVAESAADAVAAPAPETKSDAEPAAAPVPETKPVAESAADAVAAPAPETKSDAEPA
AAADPAPEIKSDAAAADPAPGTKADAAATDAAPGAEPDAAPLENTAADKGGSEESSQPVNNVNNGHST

Rice MBD1
EITVEESKEAPTTTKEATHRISRGIHDKGHSLTRKLKSDGNVDTTQILHNLHEDELAGFEESWKGNARHHLAGLNQNAGTSNNNNQVTVA
PVDVAGNPLGVGLFLEESKAVIKDGTSEDRNHVSYQSPKGFLLYIYGSKSVNCXVVESSKIQVQRILI

Arabidopsis MBD1, GI:9392683
mddgdlgnnh hnflggagnr lsaeslplid trllsqselr alsqcsslsp sssaslaasa ggdddltpki drsvfnesag
srkqtflrlr larhpqppee ppspqrqrdd ssreeqtqva sllrslfnvd snqskeeede geeeledneg qihynsyvyq
rpnldsiqnv liqgtsgnki krkrgrprki rnpseenevl dltgeastyv fvdktssnlg mvsrvgssgi sldsnsvkrk
rgrppknkee imnlekrdsa ivnisafdke elvvnlenre gtivdlsala svsedpyeee lrritvglkt keeilgfleq
lngewvnigk kkkvvnacdy ggylprgwrl mlyikrkgsn lllacrryis pdgqqfetck evstylrsll espsknqhyy
lqsdnktlgq qpvianesll gnsdsmdset mqylesgrts sevfeeakav engneadrvk tslmqkddna dflngvednd
ddmkkrdgnm enlatlsnse mtkslptttn elqqyfssqi nrvq

Table 8: Amino Acid Sequence of Gal4-VP16 Transcriptional Activator

		aagcttggatccaaca				atg	aag	ctc	ctg	tcc	tcc	atc	gag	cag	gcc	tgc	gac	atc	tgc
						M	K	L	L	S	S	I	E	Q	A	C	D	I	C
cgc	ctc	aag	aag	ctc	aag	tgc	tcc	aag	gag	aag	ccg	aag	tgc	gcc	aag	tgt	ctg	aag	aac
R	L	K	K	L	K	C	S	K	E	K	P	K	C	A	K	C	L	K	N
aac	tgg	gag	tgt	cgc	tac	tct	ccc	aaa	acc	aag	cgc	tcc	ccg	ctg	acc	cgc	gcc	cac	ctc
N	W	E	C	R	Y	S	P	K	T	K	R	S	P	L	T	R	A	H	L
acc	gaa	gtg	gag	tcc	cgc	ctg	gag	cgc	ctg	gag	cag	ctc	ttc	ctc	ctg	atc	ttc	cct	cga
T	E	V	E	S	R	L	E	R	L	E	Q	L	F	L	L	I	F	P	R
gag	gac	ctc	gac	atg	atc	ctg	aaa	atg	gac	tcc	ctc	cag	gac	atc	aaa	gcc	ctg	ctc	acc
E	D	L	D	M	I	L	K	M	D	S	L	Q	D	I	K	A	L	L	T
ggc	ctc	ttc	gtc	cag	gac	aac	gtg	aac	aaa	gac	gcc	gtc	acc	gac	cgc	ctg	gcc	tcc	gtg
G	L	F	V	Q	D	N	V	N	K	D	A	V	T	D	R	L	A	S	V
gag	acc	gac	atg	ccc	ctc	acc	ctg	cgc	cag	cac	cgc	atc	agc	gcg	acc	tcc	tcc	tcg	gag
E	T	D	M	P	L	T	L	R	Q	H	R	I	S	A	T	S	S	S	E
gag	agc	agc	aac	aag	ggc	cag	cgc	cag	ttg	acc	gtc	tcg	acg	gcc	ccc	ccg	acc	gac	gtc
E	S	S	N	K	G	Q	R	Q	L	T	V	S	T	A	P	P	T	D	V
agc	ctg	ggg	gac	gag	ctc	cac	tta	gac	ggc	gag	gac	gtg	gcg	atg	gcg	cat	gcc	gac	gcg
S	L	G	D	E	L	H	L	D	G	E	D	V	A	M	A	H	A	D	A
cta	gac	gat	ttc	gat	ctg	gac	atg	ttg	ggg	gac	ggg	gat	tcc	ccg	ggg	ccg	gga	ttt	acc
L	D	D	F	D	L	D	M	L	G	D	G	D	S	P	G	P	G	F	T
ccc	cac	gac	tcc	gcc	ccc	tac	ggc	gct	ctg	gat	atg	gcc	gac	ttc	gag	ttt	gag	cag	atg
P	H	D	S	A	P	Y	G	A	L	D	M	A	D	F	E	F	E	Q	M
ttt	acc	gat	gcc	ctt	gga	att	gac	gag	tac	ggt	ggg	tagatct							
F	T	D	A	L	G	I	D	E	Y	G	G	*							

Table 9: Amino Acid Sequence of FIE-15G-ESA1 Polypeptide

MSKITLGNESIVGSLTPSNKKSYKVTNRIQEGKKPLYAVVFNFLDARFFDVFVTAGGNRITLYNCLGDGAISALQSYADEDKEESFYTVS
WACGVNGNPYVAAGGVKGIIRVIDVNSETIHKSLVGHGDSVNEIRTQPLKPQLVITASKDESVRLWNVETGICILIFAGAGGHRYEVLSV
DFHPSDIYRFASCGMDTTIKIWSMKEFWTYVEKSFTWTDDPSKFPTKFVQFPVFTASIHTNYVDCNRWFGDFILSKSVDNEILLWEPQLK
ENSPGEGASDVLLRYPVPMCDIWFIKFSCDLHLSSVAIGNQEGKVYVWDLKSCPPVLITKLSHNQSKSVIRQTAMSVDGSTILACCEDGT
IWRWDVITKGSPGGGGGGGGGGGGGGGMRTHIEGHEELDAASLREHEEFTKVKNISTIELGKYEIETWYFSPFPPEYNDCVKLFFCEFCL
NFMKRKEQLQRHMRKCDLKHPPGDEIYRSGTLSMFEVDGKKNKVYAQNLCYLAKLFLDHKTLYYDVDLFLFYVLCECDDRGCHMVGYFSK
EKHSEEAYNLACILTLPSYQRKGYGKFLIAFSYELSKKEGKVGTPXKTLVGSRLTKLQRLLDSCSIRNLEKT

Table 10: Nucleotide Sequence of FIE-15G-ESA1

cggagtactgtcctccgagcggagtactgtcctccgagcggagtactgtcctccgagcggagtactgtcctccgagcggagtactgtcct
ccgagcggagactctagaacgattatttaggtgataagagtggacaatgatcgttgacacgtggacggtccacaaattctagttttgcct
ataagtatcaaagctgaatgtgtaagttggatccaacaccagttgttgttgcatgagagacttgtgagcttagattagtgtgcgagagtc
agacagagagagagatttcgaatatcgaatgtcgaagataaccttagggaacgagtcaatagttgggtctttgactccatcgaataagaa
atcgtacaaagtgacgaataggattcaggaagggaagaaacctttgtatgctgttgttttcaacttccttgatgctcgtttcttcgatgt
cttcgttaccgctggtggaaatcggattactctgtacaattgtctcggagatggtgccatatcagcattgcaatcctatgctgatgaaga
taaggaagagtcgttttacacggtaagttgggcgtgtggcgttaatgggaacccatatgttgcggctggaggagtaaaaggtataatccg
agtcattgacgtcaacagtgaaacgattcataagagtcttgtgggtcatggagattcagtgaacgaaatcaggacacaacctttaaaacc
tcaacttgtgattactgctagcaaggatgaatctgttcgtttgtggaatgttgaaactgggatatgtattttgatatttgctggagctgg
aggtcatcgctatgaagttctaagtgtggattttcatccgtctgatatttaccgctttgctagttgtggtatggacaccactattaaaat
atggtcaatgaaagagttttggacgtacgtcgagaagtcattcacatggactgatgatccatcaaaattccccacaaaatttgtccaatt
ccctgtatttacagcttccattcatacaaattatgtagattgtaaccgttggtttggtgattttatcctctcaaagagtgtggacaacga
gatcctgttgtgggaaccacaactgaaagagaattctcctggcgagggagcttcagatgttctattaagatacccggttccaatgtgtga
tatttggtttatcaagttttcttgtgacctccatttaagttctgttgcgataggtaatcaggaaggaaaggtttatgtctgggatttgaa
aagttgccctcctgttttgattacaaagttatcacacaatcaatcaaagtctgtaatcaggcaaacagccatgtctgtcgatggaagcac
gattcttgcttgctgcgaggacgggactatatggcgctgggacgtgattaccaagggatcccccggaggtggaggtggaggtggaggtgg
aggtggaggtggaggtggaatgaggacacatatagagggtcatgaagagctggatgcagcaagtttgcgtgaacatgaagagttcacgaa
agtgaagaacatatcaacaattgagcttggaaaatatgagattgagacttggtacttctccccttttccgccagaatacaatgactgtgt
gaagctctttttttgtgagttttgcctgaacttcatgaaacgcaaagagcagcttcaaaggcatatgagraagtgtgacctgaagcaccc
acctggtgatgaaatttaccgaagtggtaccttgtcaatgtttgaggtagatggcaaaaagaacaaggtttatgcacagaatctctgcta
cctggcaaagttatttcttgaccacaaaactctttactacgatgttgatttgtttctattctacgttctttgcgaatgtgatgaccgagg
atgccacatggttgggtacttttcaaaggagaagcattcggaagaagcatacaacttagcttgcattctaaccctgccttcatatcaaag
aaaaggctatggaaagttcttaatagccttttcctatgaactgtcaaagaaagagggaaaagttgggacaccggraaagacccttgtcgg
atctaggcttactaagctacagaggttattggactcgtgttctattagaaatcttgaaaaaacataactcgagggggggcccgctagagt
cctgctttaatgagatatgcgagacgcctatgatcgcatgatatttgctttcaattctgttgtgcacgttgtaaaaaacctgagcatgtg
tagctcagatccttaccgccggtttcggttcattctaatgaatatatcacccgttactatcgtatttttatgaataatattctccgttca
atttactgattgtaccctactacttatatgtacaatattaaaatgaaaacaatatattgtgctgaataggtttatagcgacatctatgat
agagcgccacaataacaaacaattgcgttttattattacaaatccaattttaaaaaaagcggcagaaccggtcaaacctaaaagactgat
tacataaatcttattcaaatttcaaaaggccccaggggctagtatctacgacacaccgagcggcgaactaataacgttcactgaagggaa
ctccggttccccgccggcgcgcatgggtgagattccttgaagttgagtattggccgtccgctctaccgaaagttacgggcaccattcaac
ccggtccagcacggcggccgggtaaccgacttgctgccccgagaattatgcagcatttttttggtgtatgtgggccccaaatgaagtgca
ggtcaaaccttgacagtgacg caaatcgttgggcgggtccagggcgaattttgcgacaacatgtcgaggctcagcag

Table 11: Amino Acid Sequence of MEA-15G-ESA1 Polypeptide

MEKENHEDDGEGLPPELNQIKEQIEKERFLHIKRKFELRYIPSVATHASHHQSFDLNQPAAEDDNGGDNKSLLSRMQNPLRHFSASSDYN
SYEDQGYVLDEDQDYALEEDVPLFLDEDVPLLPSVKLPIVEKLPRSITWVFTKSSQLMAESDSVIGKRQIYYLNGEALELSSEEDEEDEE
EDEEEIKKEKCEFSEDVDRFIWTVGQDYGLDDLVVRRALAKYLEVDVSDILERYNELKLKNDGTAGEASDLTSKTITTAFQDFADRRHCR
RCMIFDCHMHEKYEPESRSSEDKSSLFEDEDRQPCSEHCYLKVRSVTEADHVMDNDNSISNKIVVSDPNNTMWTPVEKDLYLKGIEIFGR
NSCDVALNILRGLKTCLEIYNYMREQDQCTMSLDLNKTTQRHNQVTKKVSRKSSRSVRKKSRLRKYARYPPALKKTTSGEAKFYKHYTPC
TCKSKCGQQCPCLTHENCCEKYCGCSKDCNNRFGGCNCAIGQCTNRQCPCFAANRECDPDLCRSCPLSCGDGTLGETPVQIQCKNMQFLL
QTNKKILIGKSDVHGWGAFTWDSLKKNEYLGEYTGELITHDEANERGRIEDRIGSSYLFTLNDQLEIDARRKGNEFKFLNHSARPNCYAK
LMIVRGDQRIGLFAERAIEEGEELFFDYCYGPEHADWSRGREPRKTGASKRSKEARPARGSPGGGGGGGGGGGGGGGMRTHIEGHEELDA
ASLREHEEFTKVKNISTIELGKYEIETWYFSPFPPEYNDCVKLFFCEFCLNFMKRKEQLQRHMRKCDLKHPPGDEIYRSGTLSMFEVDGK
KNKVYAQNLCYLAKLFLDHKTLYYDVDLFLFYVLCECDDRGCHMVGYFSKEKHSEEAYNLACILTLPSYQRKGYGKFLIAFSYELSKKEG
KVGTPXKTLVGSRLTKLQRLLDSCSIRNLEKT

Table 12: Nucleotide Sequence of MEA-15G-ESA1

atggagaaggaaaaccatgaggacgatggtgagggtttgccacccgaactaaatcagataaaagagcaaatcgaaaaggagagatttctg
catatcaagagaaaattcgagctgagatacattccaagtgtggctactcatgcttcacaccatcaatcgtttgacttaaaccagcccgct
gcagaggatgataatggaggagacaacaaatcacttttgtcgagaatgcaaaacccacttcgtcatttcagtgcctcatctgattataat
tcttacgaagatcaaggttatgttcttgatgaggatcaagattatgctcttgaagaagatgtaccattatttcttgatgaagatgtacca
ttattaccaagtgtcaagcttccaattgttgagaagctaccacgatccattacatgggtcttcaccaaaagtagccagctgatggctgaa
agtgattctgtgattggtaagagacaaatctattatttgaatggtgaggcactagaattgagcagtgaagaagatgaggaagatgaagaa
gaagatgaggaagaaatcaagaaagaaaaatgcgaattttctgaagatgtagaccgatttatatggacggttgggcaggactatggtttg
gatgatctggtcgtgcggcgtgctctcgccaagtacctcgaagtggatgtttcggacatattggaaagatacaatgaactcaagcttaag
aatgatggaactgctggtgaggcttctgatttgacatccaagacaataactactgctttccaggattttgctgatagacgtcattgccgt
cgttgcatgatattcgattgtcatatgcatgagaagtatgagcccgagtctagatccagcgaagacaaatctagtttgtttgaggatgaa
gatagacaaccatgcagtgagcattgttacctcaaggtcaggagtgtgacagaagctgatcatgtgatggataatgataactctatatca
aacaagattgtggtctcagatccaaacaacactatgtggacgcctgtagagaaggatctttacttgaaaggaattgagatatttgggaga
aacagttgtgatgttgcattaaacatacttcgggggcttaagacgtgcctagagatttacaattacatgcgcgaacaagatcaatgtact
atgtcattagaccttaacaaaactacacaaagacacaatcaggttaccaaaaagtatctcgaaaaagtagtaggtcggtccgcaaaaaa
tcgagactccgaaaatatgctcgttatccgcctgctttaaagaaaacaactagtggagaagctaagttttataagcactacacaccatgc
acttgcaagtcaaaatgtggacagcaatgcccttgtttaactcacgaaaattgctgcgagaaatattgcgggtgctcaaaggattgcaac
aatcgctttggaggatgtaattgtgcaattggccaatgcacaaatcgacaatgtccttgttttgctgctaatcgtgaatgcgatccagat
ctttgtcggagttgtcctcttagctgtggagatggcactcttggtgagacaccagtgcaaatccaatgcaagaacatgcaattcctcctt
caaaccaataaaaagattctcattggaaagtctgatgttcatggatggggtgcatttacatgggactctcttaaaaagaatgagtatctc
ggagaatatactggagaactgatcactcatgatgaagctaatgagcgtgggagaatagaagatcggattggttcttcctacctctttacc
ttgaatgatcagctcgaaatcgatgctcgccgtaaaggaaacgagttcaaatttctcaatcactcagcaagacctaactgctacgccaag
ttgatgattgtgagaggagatcagaggattggtctatttgcggagagagcaatcgaagaaggtgaggagcttttcttcgactactgctat
ggaccagaacatgcggattggtcgcgtggtcgagaacctagaaagactggtgcttctaaaaggtctaaggaagcccgtccagctcgtgga
tcccccggaggtggaggtggaggtggaggtggaggtggaggtggaatgaggacacatatagagggtcatgaagagctggatgca
gcaagtttgcgtgaacatgaagagttcacgaaagtgaagaacatatcaacaattgagcttggaaaatatgagattgagacttggtacttc
tccccttttccgccagaatacaatgactgtgtgaagctctttttttgtgagttttgcctgaacttcatgaaacgcaaagagcagcttcaa
aggcatatgagraagtgtgacctgaagcacccacctggtgatgaaatttaccgaagtggtaccttgtcaatgtttgaggtagatggcaaa
aagaacaaggtttatgcacagaatctctgctacctggcaaagttatttcttgaccacaaaactctttactacgatgttgatttgtttcta
ttctacgttctttgcgaatgtgatgaccgaggatgccacatggttgggtacttttcaaaggagaagcattcggaagaagcatacaactta
gcttgcattctaaccctgccttcatatcaaagaaaaggctatggaaagttcttaatagccttttcctatgaactgtcaaagaaagaggga
aaagttgggacaccggraaagacccttgtcggatctaggcttactaagctacagaggttattggactcgtgttctattagaaatcttgaa
aaaacataa

Table 13

Alignment of Identified and Putative Histone Acetyltransferases

GCN5 family						
hP/CAF	98	LALIKDGRVIGGICF	(6)	GFTEIVFCAVTSNEQVKGYGTHLMNHLKEYHIKHD	(4)	LTYADEYAIGYF-KKQGFSKEI
hGCN5	546	LALIKDGRVIGGICF	(6)	GFTEIVFCAVTSNEQVKGYGTHLMNHLKEYHIKHN	(4)	LTYADEYAIGYFIKKQGFSKDI
yGcn5p	141	AVIRKPLTVVGGITY	(6)	EFAEIVFCAISSTEQVRGYGAHLMNHLKDYVRNTS	(4)	LTYADNYAIGYF-KKQGFTKEI
tHatA	149	VILKNKQKVIGGICF	(6)	RFAEVAFLAVTANEQVRGYGTRLMNKFKDHMQKQN	(5)	LTYADNFAIGYF-KKQGFTKEH
Consensus		lalikdgrViGGIcf		gF EivFcAvtsnEQVkGYGthLMNhlKeyhikhn		LTYAD yAIGYF KKQGFsKei
SAS family (aka MYST)						
ySas2		--STKPMGFFSKDLVSY		QQNNLACILIFPPYQRRGLGLLLIEFSYKLSQLEGV		ISGPEVPLSPFGLIGYLK
hTIP60		CKGEHIVGYFSKEKESТ		EDYNVACILTLPPYQRRGYGKLLIEFSYELSKVEGK		TGTPEKPLSDLGLLSYRS
dmMOF		-EGSHIVGYFSKEKKSL		ENYNVACILVLPPHQRKGFGKLLIAFSYELSRKEGV		IGSPEKPLSDLGRLSYRS
yMof		ELGEHILVGYFSKEKESA		DGYNVACILTLPQYQRMGYGKLLIEFSYELSKKENK		VGSPEKPLSDLGLLSYRA
spYD6D		GQECKVIGYFSKEKRSA		SDYNVSCILTLPIYQRRGYGVFLIDFSYLLTQVEGK		LGSPEKPLSDLGLVTYRS
hMOZ		VKGCHLVGYFSKEKHCQ		QKYNVSCIMILPQYQRKGYGRFLIDFSYLLSKREGQ		AGSPEKPLSDLGRLSYMA
ySas3		AAKEHFVGYFSKEKFNS		NDYNLSCILTLPIYQRKGYGQFLMEFSYLLSRKESK		FGTPEKPLSDLGLLTYRT
ceR07B5.8		DIGCHFAGYFSKEKYEP		DVNNLSCIMTLPCYQEMGLGRFLIDISYALSRKEKW		FGGPEQPLSELGRKAYGG
Consensus		g hivGyFSKek s		yNv CIltlP yQrrGyGk LiefSY LsKkEg		gtPEkPLSdlGllsYr
yKL087	448	YEPDKKDILTGLLRL	(27)	GSVVPLHSRDPRKFCHQGFGTLLMEEAERIAKEEH	(5)	SVISGVGVRNYYGKLGYELDG
yHat1p	179	LLNKKTKELIGFVTT	(19)	FRAKISQFLIFPPYQNKGHGSCLYEAIIQSWLEDK	(14)	DDLRDRNDIQRLRKLGYDAVF
YOR109		HFCYELKKLLSNGSF		YSSD---SDLTSTLQHRGYGQHSLST--DTYEEEY		
YPR193c				--IYINDLYVDENSRVKGAGGKLTQEVYDEADKLG		T
YEH6	32	AFDTETGDAIGFA-HY	(10)	EVVYMNDLYVTERARVKGVGRKLIEEVYSRADELG	(7)	TDHYNHRAQLLYTKVAYKTDK
YMR069w		MLTEETGLVEGDALHE	(3)	PVIYLYEVHVASAHRGHGIGRRLLEHALCDG		
ySpt10				VVTGTFLVN-AGI-RGKGIGKTLMETFIEWSKKLG		
spATS1				VGIYLEDLYVRPQFRGKGYGSYLLSYMARESLRI		
Concensus				-viyi dlyv RgkGvGrkLiefv ea klg		

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

What is claimed is:

1. A chimeric polypeptide comprising:
   a. a first polypeptide segment that exhibits histone acetyltransferase activity; and
   b. a second polypeptide segment, wherein said second polypeptide segment has 40% or greater sequence identity to a subunit of a histone deacetylase chromatin-associated protein complex,

   wherein a terminus of said second polypeptide segment is linked to a terminus of said first polypeptide segment via at least one covalent bond.

2. The chimeric polypeptide of claim 1, wherein said second polypeptide has 60% or greater sequence identity to a subunit of a histone deacetylase chromatin-associated protein complex.

3. The chimeric polypeptide of claim 1, wherein said second polypeptide has 80% or greater sequence identity to a subunit of a histone deacetylase chromatin-associated protein complex.

4. The chimeric polypeptide of claim 1, wherein said second polypeptide has 90% or greater sequence identity to a subunit of a histone deacetylase chromatin-associated protein complex.

5. The chimeric polypeptide of claim 1, wherein said subunit exhibits scaffold activity.

6. The chimeric polypeptide of claim 1, wherein said subunit exhibits DNA binding activity.

7. The chimeric polypeptide of claim 1, wherein said subunit exhibits ATPase-dependent helicase activity.

8. The chimeric polypeptide of claim 1, wherein said subunit exhibits histone deacetylase activity.

9. The chimeric polypeptide of claim 1, wherein said first and said second polypeptide segments are directly linked via a peptide bond.

10. The chimeric polypeptide of claim 9, wherein the C-terminal amino acid of said first polypeptide segment is linked to the N-terminal amino acid of said second polypeptide segment.

11. The chimeric polypeptide of claim 9, wherein the N-terminal amino acid of said first polypeptide segment is linked to the C-terminal amino acid of said second polypeptide segment.

12. The chimeric polypeptide of claim 1, wherein said first and said second polypeptide segments are indirectly linked via one or more intervening amino acids that are situated between said first and said second polypeptide segments.

13. The chimeric polypeptide of claim 12, wherein the C-terminal amino acid of said first polypeptide segment is linked to one of said one or more intervening amino acids, and wherein the N-terminal amino acid of said second polypeptide segment is linked to one of said one or more intervening amino acids.

14. The chimeric polypeptide of claim 12, wherein the N-terminal amino acid of said first polypeptide segment is linked to one of said one or more intervening amino acids, and wherein the C-terminal amino acid of said second polypeptide segment is linked to one of said one or more intervening amino acids.

15. The chimeric polypeptide of claim 12, wherein said first and said second polypeptide segments are indirectly linked via 1 to 50 intervening amino acids.

16. The chimeric polypeptide of claim 15, wherein said first and said second polypeptide segments are indirectly linked via 10 to 50 intervening amino acids.

17. The chimeric polypeptide of claim 15, wherein said intervening amino acids comprise at least one alanine residue.

18. The chimeric polypeptide of claim 15, wherein said intervening amino acids comprise at least one glycine residue.

19. A nucleic acid construct encoding the polypeptide of claim 1.

20. A eukaryotic organism comprising the chimeric polypeptide of claim 1.

21. A eukaryotic organism comprising a nucleic acid encoding a chimeric polypeptide, said chimeric polypeptide comprising:
   a. a first polypeptide segment that exhibits histone acetyltransferase activity; and
   b. a second polypeptide segment, wherein said second polypeptide segment has 40% or greater sequence identity to a subunit of a histone deacetylase chromatin-associated protein complex,

   wherein a terminus of said second polypeptide segment is covalently linked to a terminus of said first polypeptide segment.

22. A eukaryotic organism comprising:
   a. a first nucleic acid construct comprising a first promoter and a transcription activator element operably linked to a coding sequence, said coding sequence encoding:
      i) a first polypeptide segment that exhibits histone acetyltransferase activity; and
      ii) a second polypeptide segment, wherein said second polypeptide segment has 40% or greater sequence identity to a subunit of a histone deacetylase chromatin-associated protein complex,

wherein a terminus of said second polypeptide segment is covalently linked to a terminus of said first polypeptide segment; and

b. a second nucleic acid construct comprising a second promoter conferring cell type-specific transcription, said second promoter operably linked to a coding sequence for a polypeptide that binds said transcription activator element.

23. The eukaryotic organism of claim 22, wherein said eukaryotic organism is an animal.

24. The eukaryotic organism of claim 22, wherein said eukaryotic organism is a plant.

25. The eukaryotic organism of claim 24, wherein said plant is a monocot.

26. The eukaryotic organism of claim 25, wherein said monocot is selected from the group consisting of corn and rice.

27. The eukaryotic organism of claim 24, wherein said plant is a dicot.

28. The eukaryotic organism of claim 27, wherein said dicot is selected from the group consisting of soybean and rape.

29. The eukaryotic organism of claim 24, wherein said plant comprises an agent or mutation that alters the DNA methylation state in said plant relative to a corresponding plant that lacks said agent or mutation.

30. The eukaryotic organism of claim 29, wherein said DNA methylation state is decreased relative to a corresponding organism that lacks said agent or mutation.

31. The eukaryotic organism of claim 30, wherein said mutation is in a C5 DNA methyltransferase gene.

32. The eukaryotic organism of claim 30, wherein said agent affects expression of a C5 DNA methyltransferase gene.

33. The eukaryotic organism of claim 32, wherein said agent is an antisense nucleic acid.

34. A method for detecting the expression of one or more genes in a eukaryote, said method comprising:

a. isolating macromolecules from one or more specific cells in said eukaryote, said eukaryote comprising a nucleic acid construct that has a promoter operably linked to a coding sequence that encodes:

i) a first polypeptide segment that exhibits histone acetyltransferase activity; and

ii) a second polypeptide segment, wherein said second polypeptide segment has 40% or greater sequence identity to a subunit of a histone deacetylase chromatin-associated protein complex,

wherein a terminus of said second polypeptide segment is covalently linked to a terminus of said first polypeptide segment; and

b. determining the presence or amount of at least one of said macromolecules in at least one of said specific cells.

35. The method of claim 34, wherein said macromolecules are polypeptides.

36. The method of claim 34, wherein said macromolecules are nucleic acids.

37. The method of claim 34, wherein said eukaryote is an animal.

38. The method of claim 34, wherein said eukaryote is a plant.

39. The method of claim 38, wherein said promoter confers cell-type specific transcription in a plant reproductive tissue.

40. The method of claim 39, wherein said reproductive tissue is a flower.

41. The method of claim 40, wherein said reproductive tissue is selected from the group consisting of: ovule, and central cell.

42. The method of claim 39, wherein said reproductive tissue is a seed.

43. The method of claim 42, wherein said reproductive tissue is selected from the group consisting of: endosperm, embryo, and zygote.

44. The method of claim 38, wherein said promoter confers cell-type specific transcription in a plant vegetative tissue.

45. A method for detecting the expression of one or more genes in a eukaryote, said method comprising:

   a. isolating macromolecules from one or more specific cells in said eukaryote, said eukaryote comprising:

      i) a first nucleic acid construct having a first promoter and a transcription activator element operably linked to a coding sequence, said coding sequence encoding:

         a) a first polypeptide segment that exhibits histone acetyltransferase activity; and

         b) a second polypeptide segment, wherein said second polypeptide segment has 40% or greater sequence identity to a subunit of a histone deacetylase chromatin-associated protein complex,

wherein a terminus of said second polypeptide segment is covalently linked to a terminus of said first polypeptide segment; and

ii) a second nucleic acid construct comprising a second promoter conferring cell type-specific transcription, said second promoter operably linked to a coding sequence for a polypeptide that binds said transcription activator element; and

b. determining the presence or amount of at least one of said macromolecules in at least one of said specific cells.

46. The method of claim 45, wherein said macromolecules are polypeptides.

47. The method of claim 45, wherein said macromolecules are nucleic acids.

48. The method of claim 45, wherein said eukaryote is an animal.

49. The method of claim 45, wherein said eukaryote is a plant.

50. The method of claim 49, wherein said second promoter confers cell-type specific transcription in a plant reproductive tissue.

51. The method of claim 50, wherein said reproductive tissue is a flower.

52. The method of claim 51, wherein said reproductive tissue is selected from the group consisting of: ovule, and central cell.

53. The method of claim 50, wherein said reproductive tissue is a seed.

54. The method of claim 53, wherein said reproductive tissue is selected from the group consisting of: endosperm, embryo, and zygote.

55. The method of claim 49, wherein said second promoter confers cell-type specific transcription in a plant vegetative tissue.

56. A method for modulating gene expression in a eukaryote, said method comprising making a eukaryote having a nucleic acid construct comprising a cell-type specific promoter operably linked to a coding sequence that encodes:

a. a first polypeptide segment that exhibits histone acetyltransferase activity; and

b. a second polypeptide segment, wherein said second polypeptide segment has 40% or greater sequence identity to a subunit of a histone deacetylase chromatin-associated protein complex,

wherein a terminus of said second polypeptide segment is covalently linked to a terminus of said first polypeptide segment; and

wherein said eukaryote exhibits modulated gene expression in cells in which said promoter confers cell-type specific transcription.

57. The method of claim 56, wherein said eukaryote has compositional alterations relative to a corresponding organism that lacks said nucleic acid construct.

58. The method of claim 56, wherein said eukaryote has developmental alterations relative to a corresponding organism that lacks said nucleic acid construct.

59. The method of claim 56, wherein said eukaryote has phenotypic alterations relative to a corresponding organism that lacks said nucleic acid construct.

60. The method of claim 56, wherein said eukaryote is an animal.

61. The method of claim 56, wherein said eukaryote is a plant.

62. The method of claim 61, wherein said promoter confers cell-type specific transcription in a plant reproductive tissue.

63. The method of claim 62, wherein said reproductive tissue is a flower.

64. The method of claim 63, wherein said reproductive tissue is selected from the group consisting of: ovule, and central cell.

65. The method of claim 62, wherein said reproductive tissue is a seed.

66. The method of claim 65, wherein said reproductive tissue is selected from the group consisting of: endosperm, embryo, and zygote.

67. The method of claim 61, wherein said promoter confers cell-type specific transcription in a plant vegetative tissue.

68. The method of claim 61, wherein said plant comprises an agent or mutation that alters the DNA methylation state in said plant relative to a corresponding plant that lacks said agent or mutation.

69. The method of claim 68, wherein said DNA methylation state is decreased relative to a corresponding plant that lacks said agent or mutation.

70. The method of claim 69, wherein said mutation is in a C5 DNA methyltransferase gene.

71. The method of claim 69, wherein said agent affects expression of a C5 DNA methyltransferase gene.

72. The method of claim 71, wherein said agent is an antisense nucleic acid.

73. The method of claim 61, wherein said modulated gene expression alters seed development.

74. The method of claim 61, wherein said modulated gene expression alters embryo development.

75. The method of claim 61, wherein said modulated gene expression alters endosperm development.

76. The method of claim 61, wherein said modulated gene expression alters seed yield by mass.

77. A method for modulating gene expression in a eukaryote, said method comprising making a eukaryote, said eukaryote comprising:

a. a first nucleic acid construct having a first promoter and a transcription activator element operably linked to a coding sequence, said coding sequence encoding:

i) a first polypeptide segment that exhibits histone acetyltransferase activity; and

ii) a second polypeptide segment, wherein said second polypeptide segment has 40% or greater sequence identity to a subunit of a histone deacetylase chromatin-associated protein complex,

wherein a terminus of said second polypeptide segment is covalently linked to a terminus of said first polypeptide segment; and

b) a second nucleic acid construct comprising a second promoter conferring cell type-specific transcription, said second promoter operably linked to a coding sequence for a polypeptide that binds said transcription activator element,

wherein said eukaryote exhibits modulated gene expression in cells in which said second promoter confers cell-type specific expression.

78. The method of claim 77, wherein said eukaryote has compositional alterations relative to a corresponding organism that lacks said nucleic acid construct.

79. The method of claim 77, wherein said eukaryote has developmental alterations relative to a corresponding organism that lacks said nucleic acid construct.

80. The method of claim 77, wherein said eukaryote has phenotypic alterations relative to a corresponding organism that lacks said nucleic acid construct.

81. The method of claim 77, wherein said eukaryote is an animal.

82. The method of claim 77, wherein said eukaryote is a plant.

83. The method of claim 82, wherein said promoter confers cell-type specific transcription in a plant reproductive tissue.

84. The method of claim 83, wherein said reproductive tissue is a flower.

85. The method of claim 84, wherein said reproductive tissue is selected from the group consisting of: ovule, and central cell.

86. The method of claim 83, wherein said reproductive tissue is a seed.

87. The method of claim 86, wherein said reproductive tissue is selected from the group consisting of: endosperm, embryo, and zygote.

88. The method of claim 82, wherein said promoter confers cell-type specific transcription in a plant vegetative tissue.

89. The method of claim 82, wherein said plant comprises an agent or mutation that alters the DNA methylation state in said plant relative to a corresponding plant that lacks said agent or mutation.

90. The method of claim 89, wherein said DNA methylation state is decreased relative to a corresponding plant that lacks said agent or mutation.

91. The method of claim 90, wherein said mutation is in a C5 DNA methyltransferase gene.

92. The method of claim 90, wherein said agent affects expression of a C5 DNA methyltransferase gene.

93. The method of claim 92, wherein said agent is an antisense nucleic acid.

94. The method of claim 82, wherein said modulated gene expression alters seed development.

95. The method of claim 82, wherein said modulated gene expression alters embryo development.

96. The method of claim 82, wherein said modulated gene expression alters endosperm development.

97. The method of claim 82, wherein said modulated gene expression alters seed yield by mass.

98. A method for making a genetically modified eukaryote, said method comprising:

a. providing a first eukaryote comprising a first nucleic acid construct, said first nucleic acid construct comprising a first promoter and a transcription activator element operably linked to a coding sequence, said coding sequence encoding:

i) a first polypeptide segment that exhibits histone acetyltransferase activity; and

ii) a second polypeptide segment, wherein said second polypeptide segment has 40% or greater sequence identity to a subunit of a histone deacetylase chromatin-associated protein complex,

wherein a terminus of said second polypeptide segment is covalently linked to a terminus of said first polypeptide segment; and

b. providing a second eukaryote comprising a second nucleic acid construct, said second nucleic acid construct comprising a second promoter conferring embryo-specific transcription, said second promoter operably linked to a coding sequence for a polypeptide that binds said transcription activator element,

c. crossing said first eukaryote and said second eukaryote to form genetically modified progeny that are sterile.

99. The method of claim 98, wherein said eukaryote is an animal.

100. The method of claim 98, wherein said eukaryote is a plant.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/19750

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :Please See Extra Sheet.
US CL : 530/350; 800/13, 3; 435, 7.2, 4, 7.1, 455
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/350; 800/13, 3; 435, 7.2, 4, 7.1, 455

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WEST, MEDLINE, BIOSIS, EMBASE, SCISEARCH, CAPLUS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MARTINEZ-BALBAS, M.A. et al. Drosophila NURF-55, a WD repeat protein involved in histone metabolism. PNAS. January 1998, Vol. 95, No. 1, pages 132-137, see entire document.	1-100
Y	US 6,166192 A (SPIEGELMAN et al) 26 December 2000, see entire document.	1-100
Y	US 6,248,520 B1 (ROEDER et al) 19 June 2001, see entire document.	1-100

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

- * Special categories of cited documents:
- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed
- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
08 SEPTEMBER 2002	02 OCT 2002

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	PETER PARAS JR
Facsimile No. (703) 305-3230	Telephone No. (703) 308-8340

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/19750

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (7):

C07K 1/00, 14/00, 17/00; A01K 67/00, 67/033; G01N 33/00, 33/53, 33/567; C12Q 1/00; C12N 15/63, 15/85, 15/87